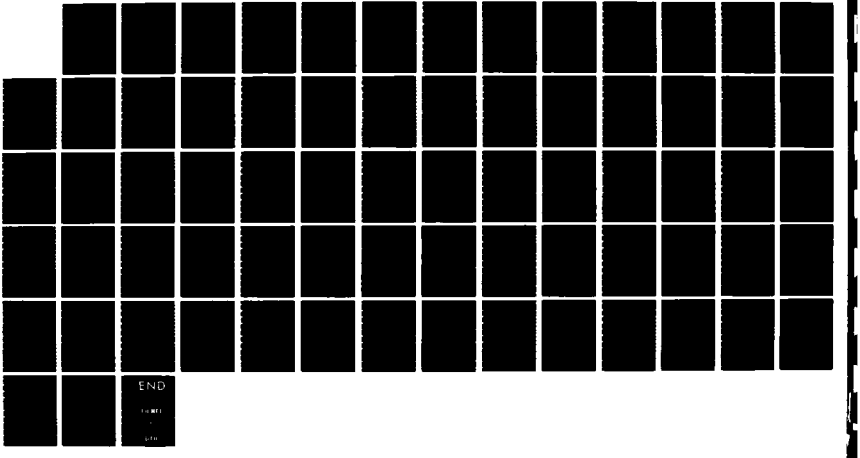
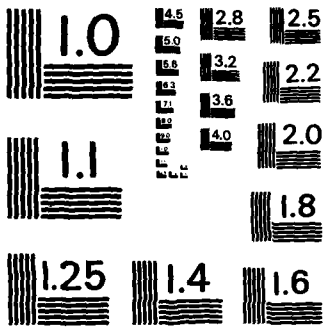
AD-A163 024 THE STRUCTURES OF DISCOURSE STRUCTURE(U) BBN LABS INC 1/1
CAMBRIDGE MA B J GROSZ ET AL NOV 85 BBN-6097
N00014-85-C-0079

UNCLASSIFIED F/G 5/7 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

BBN Laboratories Incorporated

A Subsidiary of Bolt Beranek and Newman Inc.

12

AD-A163 024

Report No. 6097

The Structures of Discourse Structure

Technical Report

November 1985

Prepared for:
Defense Advanced Research Projects Agency

DTIC
ELECTE
JAN 0 8 1986
S D

DTIC FILE COPY

This document has been approved for public release and sale; its distribution is unlimited.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER BBN Report No. 6097	2. GOVT ACCESSION NO. AD-A163 024	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE STRUCTURES OF DISCOURSE STRUCTURE		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER BBN Report No. 6097
7. AUTHOR(s) B. J. Grosz and C. L. Sidner		8. CONTRACT OR GRANT NUMBER(s) N00014-85-C-0079
9. PERFORMING ORGANIZATION NAME AND ADDRESS BBN Laboratories Inc. 10 Moulton Street Cambridge, MA 02238		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, VA 22217		12. REPORT DATE November 1985 13. NUMBER OF PAGES 66
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Natural language processing, artificial intelligence, reference, anaphora, discourse, speaker intention, focusing, dialogue.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper reports on the development of a computational theory of discourse structure. The theory is based on the thesis that discourse structure is a composite of three structures: the structure of the sequence of utterances, the structure of intentions conveyed, and the attentional state. The distinction among these components is essential to provide adequate explanations of such discourse phenomena as clue words, referring expressions and interruptions. We illustrate the —> cont'd.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (cont'd.)

application of this theory to such phenomena and discuss the recognition of intentions in discourse. Keywords: natural language processing; artificial intelligence; linguistics; speech.

Report No. 6097

THE STRUCTURES OF DISCOURSE STRUCTURE

Technical Report

B. J. Grosz
SRI International and
Center for the Study of Language and Information
and
C. L. Sidner
BBN Laboratories Inc.

November 1985

Accession For
NTIS CRA&I [X]
DTIC TAB [ ]
Unannounced [ ]
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

QUALITY INSPECTED 3

Prepared for:

Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

This report will be published in *Computational Linguistics*, 1986, and will also be published as technical note CSLI-85-39 from the Center for the Study of Language and Information, Stanford University, and Technical Note #369, from the Artificial Intelligence Center, SRI International.

TABLE OF CONTENTS

1	Introduction	1
2	The Basic Theory	3
	2.1 Linguistic Structure	4
	2.2 Intentional Structure	6
	2.3 Attentional State	9
3	Two Examples	14
	3.1 An Argument	15
	3.2 A Task-Oriented Dialogue	19
4	Some Properties of Discourse-Level Intentions	23
	4.1 The Basic Generalization	25
	4.2 The Belief Case	27
	4.3 The Action Case	29
5	Recognition Issues	31
	5.1 What Must Be Recognized	31
	5.2 Information Constraining the DSP	33
	5.3 When Is the Intention Recognized?	37
6	Application of the Theory: Interruptions	39
	6.1 Preliminary Definitions	39
	6.2 Type 1: True Interruptions	41
	6.3 Type 2: Flashbacks and Filling in Missing Pieces	43
	6.4 Type 3: Digressions	46
	6.5 Noninterruptions--"Semantic Returns"	47
7	Application of the Theory: Clue Words	48
8	Application of the Theory: Referring Expressions	53
9	Conclusions and Future Research	56

LIST OF FIGURES

Figure 1:	Discourse Segments, Focus Spaces and Dominance Hierarchy	11
Figure 2:	The Movies Essay	16
Figure 3:	Primary Intentions of the DSPs for Movies Essay	17
Figure 4:	Dominance Relationships for the DSPs of the Movies Essay	18
Figure 5:	A Segment of a Task-Oriented Dialogue	20
Figure 6:	Intentional Structure for the Task-Oriented Dialogue Segment	21
Figure 7:	The Structures of a True Interruption	42
Figure 8:	The Auxiliary Stack Model for Flashbacks	45
Figure 9:	An Initial Discourse Structure Configuration	49
Figure 10:	Attentional and Intentional Structures for a New Subsegment	50
Figure 11:	Focus Stack Transitions Leading up to Utterance (25)	54

1 Introduction

This paper presents the basic elements of a computational theory of discourse structure that simplifies and expands upon previous work. It is concerned with answers to two rather simple questions: What is discourse? What is discourse structure? As we develop it, the theory of discourse will be seen to be intimately connected with two nonlinguistic notions, namely intention and attention. Attention is an essential factor in explicating the processing of utterances in discourse. Intentions play a primary role in explaining discourse structure, defining discourse coherence, and providing a coherent conceptualization of the term "discourse" itself.

The theory is a further development and integration of two lines of research: work on focusing in discourse [16, 17, 18] and more recent work on intention recognition in discourse [39, 41, 43, 3]. Our goal has been to generalize these constructs properly to a wide range of discourse types. Grosz [op.cit.] demonstrated that the notions of focusing and task structure are necessary for understanding and producing task-oriented dialogue. One of the main generalizations of previous work will be to show that discourses are generally in some sense "task-oriented," but the kinds of "tasks" that can be engaged in are quite varied--some are physical, others mental, others linguistic. Consequently, the term "task" is unfortunate, and we will use the more general terminology of intentions (e.g., speaking of discourse purposes) for most of what we say.

Our main thesis is that the structure of any discourse is a composite of three distinct but interacting constituents: (1) the structure of the actual sequence of utterances in the discourse; (2) a structure of intentions; (3) an attentional state. The distinction among these constituents is essential to an explanation of interruptions (see Section 6), as well as to explanations of the use of certain types of referring expressions (see Section 8) and various other expressions that affect discourse segmentation and structure (see Section 7). Most related work on discourse structure (including [33, 25, 24, 9]) fails to distinguish among some (or all) of these constituents. As a result, significant generalizations are lost, and the computational mechanisms proposed are more complex than necessary. By carefully distinguishing these constituents, we are able to account for significant observations in this related work while simplifying the explanations given and computational mechanisms used.

In addition to explaining these linguistic phenomena, the theory provides an overall framework within which to answer questions about the relevance of various segments of discourse to one another and to the overall purposes of the discourse participants. Various properties of the intentional component have implications for work in general in natural-language processing. In particular, the range of intentions that underlie discourse is so varied that approaches to discourse coherence based on selecting discourse relationships from a fixed set of alternative rhetorical patterns (e.g., [22, 27, 32]) are unlikely to suffice. The intentional structure that is introduced in this paper depends instead on a small number of structural relations that can hold between intentions. This study also reveals several problems that must be confronted in expanding speech-act-related theories (e.g., [2, 8, 3]) from coverage of individual utterances to coverage of extended sequences of utterances in discourse.

Although a definition of "discourse" must await further development of the theory presented in the remainder of this paper, some properties of the phenomena we want to explain must be specified now. In particular, we take a discourse to be a piece of language behavior that typically involves multiple utterances and multiple participants. The discourse may be produced by one or more speakers (or writers) and the audience may comprise one or more hearers (or readers). Each *conversational participant (CP)* brings to the discourse a set of beliefs, goals, intentions, and other mental attitudes. These attitudes affect a CP's participation; they influence both the way utterances are produced and the way they are understood.

In the remainder of the paper, we will use the terms *initiating conversational participant (ICP)* and *other conversational participant(s) (OCP)* to distinguish the initiator of a discourse segment from other participants. In multiparty conversations of more than one segment, each participant may at different times be a speaker or a hearer. Hence these roles do not make the distinction necessary for most of the account we are providing. By speaking of ICPs and OCPs, we can highlight the purposive aspect of discourse. We will use the terms *speaker* and *hearer* only when the particular speaking/hearing activity is important for the point being made.

The next section of this paper lays out the basic theory of discourse structure and provides an overview of each of the components of discourse structure. Section 3 analyzes two sample discourses--a written text and a fragment of task-oriented

dialogue--from the perspective of the theory being developed; these two examples are used as well to illustrate various points in the remainder of the paper. Sections 4 and 5 discuss particular aspects of the intentional structure. The next three sections describe the role of the discourse structure components in explaining various properties of discourse, thereby providing evidence for the necessity of distinguishing among these components. Finally, Section 9 presents a number of outstanding research questions suggested by the theory.

2 The Basic Theory

Discourse structure is a composite of three interacting components: a linguistic structure, an intentional structure, and an attentional state. These three components of discourse structure deal with different aspects of the utterances in a discourse. Utterances--the actual saying or writing of particular sequences of phrases and clauses--are the linguistic structure's basic elements. Intentions of a particular sort and a small number of relationships between them provide the basic elements of the intentional structure. Attentional state contains information about the objects, properties, relations, and discourse intentions that are most salient at any given point. It is an abstraction of the focus of attention of the discourse participants; it serves to summarize information from previous utterances crucial for processing subsequent ones thus obviating the need for keeping a complete history of the discourse.

Together the three constituents of discourse structure supply the information needed by the CPs to determine how an individual utterance fits with the rest of the discourse--in essence, enabling them to figure why it was said and what it means. The context provided by these constituents also forms the basis for certain expectations about what is to come; these expectations too play a role in accommodating new utterances. The attentional state serves an additional purpose: namely, it furnishes the means for actually using the information in the other two structures in generating and interpreting individual utterances.

2.1 Linguistic Structure

The first component of discourse structure is the structure of the sequence of utterances that comprise a discourse.[1] Just as the words in a single sentence form constituent phrases, the utterances in a discourse are naturally aggregated into *discourse segments*. The utterances in a segment, like the words in a phrase, serve particular roles with respect to that segment. In addition, the discourse segments, like the phrases, fulfill certain functions with respect to the overall discourse. Although two consecutive utterances may be in the same discourse segment, it is also common for two consecutive utterances to be in different segments. It is also possible for two utterances that are nonconsecutive to be in the same segment.

The factoring of discourses into segments has been observed across a wide range of discourse types. Grosz [16] showed this for task-oriented dialogues. Linde [25] found it valid for descriptions of apartments; Linde and Goguen [24] describe such structuring in the Watergate transcripts. Reichman [33] observed it in informal debates, explanations, and therapeutic discourse. Cohen [9] found similar structures in essays in rhetorical texts. Polanyi and Scha [31] discuss this feature of narratives.

Although different researchers with different theories have examined a variety of discourse types and found discourse-level segmentation, there has been very little investigation of the extent of agreement about where the segment boundaries lie. There have been no psychological studies of the consistency of recognition of section boundaries. However, Mann [26] asked several people to segment a set of dialogues. He has reported [personal communication] that his subjects segmented the discourses approximately the same; their disagreements were about utterances at the boundaries

[1] The use of the phrase "linguistic structure" to refer to the structure of sequences of utterances is a natural extension of its use in traditional linguistic theories to refer to the syntactic structure of individual sentences. To avoid confusion the phrase "linguistic structure" will be used in this paper only to refer to the structure of a sequence of utterances composing a discourse or discourse segment.

of segments.[2] Several studies of spontaneously produced discourses provide additional evidence of the existence of segment boundaries, as well as suggesting some of the linguistic clues available for detecting boundaries. Chafe [6, 7] found differences in pause lengths at segment boundaries. Butterworth [5] found speech rate differences that correlated with segments; speech rate is slower at start of segment [more hesitations] than toward end.

There is a two-way interaction between the discourse segment structure and the utterances constituting the discourse: linguistic expressions can be used to convey information about the discourse structure; conversely, the discourse structure constrains the interpretation of expressions (and hence affects what a speaker says and how a hearer will interpret what is said). Not surprisingly, linguistic expressions are among the primary indicators of discourse segment boundaries. The explicit use of certain words and phrases (e.g., "in the first place") and more subtle clues, such as changes in tense and aspect, are included in the repertoire of linguistic devices that function, wholly or in part, to indicate these boundaries [33, 9, 29]. As discussed in Section 7, these linguistic boundary markers can be divided according to whether they explicitly indicate changes in the intentional structure or in the attentional state of the discourse. The differential use of these linguistic markers provides one piece of evidence for considering these two components to be distinct. In addition, because these linguistic devices function explicitly as indicators of discourse structure, it becomes clear that they are best seen as providing information at the discourse level, and not at that of the sentence; hence, certain kinds of questions (e.g., about their contribution to the truth conditions of an individual sentence) do not make sense.

Just as linguistic devices affect structure, so the discourse segmentation affects the interpretation of linguistic expressions in a discourse. Referring expressions provide the primary example of this effect.[3] The segmentation of discourse constrains

---

[2]He has also reported that the subjects did not label segments nearly so consistently. We believe this fact is related to the kinds of relations the labels were dependent upon. As discussed in Section 5, there is a difference between the intentional structure we describe and the relations that others use.

[3]Referring expressions can also be used to mark a discourse boundary. For example, novelists sometimes use pronouns to indicate a new scene in a story.

the use of referring expressions by delineating certain points at which there is a significant change in what entities (objects, properties, or relations) are being discussed. For example, there are different constraints on the use of pronouns and reduced definite-noun phrases within a segment than across segment boundaries. While discourse segmentation is obviously not the only factor governing the use of referring expressions, it is an important one.

2.2 Intentional Structure

A rather straightforward property of discourses, namely, that they (or, more accurately, those who participate in them) have an overall purpose, turns out to play a fundamental role in the theory of discourse structure. In particular, some of the purposes that underlie discourses, and their component segments, provide the means of individuating discourses and of distinguishing discourses that are coherent from those that are not. These purposes also make it possible to determine when a sequence of utterances comprises more than one discourse.

Although typically the participants in a discourse may have more than one aim in participating in the discourse (e.g., a story may entertain its listeners as well as describe an event; an argument may establish a person's brilliance as well as convince someone that a claim or allegation is true), we distinguish one of these purposes as foundational to the discourse. We will refer to it as the *discourse purpose (DP)*. From an intuitive perspective, the discourse purpose is the intention that underlies engaging in the particular discourse. This intention provides both the reason a discourse (a linguistic act), rather than some other action, is being performed and the reason the particular content of this discourse is being conveyed rather than some other information. For each of the discourse segments, we can also single out one intention--the *discourse segment purpose (DSP)*. From an intuitive standpoint, the DSP specifies how this segment contributes to achieving the overall discourse purpose. The assumption that there are single such intentions will in the end prove too strong. However, its use allows us to describe the basic theory more clearly. We must leave to future research (and a subsequent paper) the exploration and discussion of the complications that result from relaxing this assumption.

Typically, an ICP will have a number of different kinds of intentions that lead to

initiating a discourse. One kind might include intentions to speak in a certain language or to utter certain words. Another might include intentions to amuse or to impress. The kinds of intentions that can serve as discourse purposes or discourse segment purposes are distinguished from other intentions by the fact that they are intended to be recognized (cf. [2, 43]), whereas other intentions are private; that is, the recognition of the DP or DSP is *essential* to its achieving its intended effect. Discourse purposes and discourse segment purposes share this property with certain utterance-level intentions that Grice [13] uses in defining utterance meaning (see Section 4).

It is important to distinguish intentions that are intended to be recognized from other kinds of intentions that are associated with discourse. Some intention that is private and not intended to be recognized may be the primary motivation for an ICP to begin a discourse. For example, the ICP may intend to impress someone or may plan to teach someone. In neither case is the ICP's intention necessarily intended to be recognized. Quite the opposite may be true in the case of impressing, as the ICP may not want the OCP to be aware of his intention. When teaching, the ICP may not care whether the OCP knows the ICP is teaching him or her. Thus, the primary intention that motivates the ICP to engage in a discourse may be private. By contrast, the discourse segment purpose is always intended to be recognized.

DPs and DSPs are basically the same sorts of intentions. If an intention is a DP, then its satisfaction is a main purpose of the discourse, whereas if it is a DSP, then its satisfaction contributes to the satisfaction of the DP. The following are some examples of the types of intentions that could serve as DP/DSPs, followed by one particular instance of each type.

1. Intend that some agent intend to perform some physical task; intend that Ruth intend to fix the flat tire.

2. Intend that some agent believe some fact; intend that Ruth believe the campfire has started.

3. Intend that some agent believe that one fact supports another; intend that Ruth believe the smell of smoke provides evidence that the campfire is started.

4. Intend that some agent intend to identify an object (existing physical object, imaginary object, plan, event, event sequence); intend that Ruth intend to identify my bicycle.

5. Intend that some agent know some property of an object; intend that Ruth know that my bicycle has a flat tire.

We have identified two structural relations that play an important role in discourse structure: *dominance* and *satisfaction-precedence*. An action that satisfies one intention, say DSP1, may be intended to provide part of the satisfaction of another, say DSP2. When this is the case, we will say that DSP1 *contributes to* DSP2; conversely, we will say that DSP2 *dominates* DSP1 (or DSP2 *DOM* DSP1). The dominance relation invokes a partial ordering on DSPs that we will refer to as the *dominance hierarchy*. For some discourses, including task-oriented ones, the order in which the DSPs are satisfied may be significant, as well as being intended to be recognized. We will say that DSP1 *satisfaction-precedes* DSP2 (or, DSP1 *SP* DSP2) whenever DSP1 must be satisfied before DSP2.[4]

Any of the intentions on the preceding list could be either a DP or a DSP. Furthermore, a given instance of any one of them could contribute to another, or to a different instance of the same type. For example, the intention that someone intend to identify some object might dominate several intentions that she or he know some property of that object; likewise, the intention to get someone to believe some fact might dominate a number of contributing intentions that that person believe other facts.

As the above list makes clear, the range of intentions that can serve as discourse, or discourse segment, purposes is open-ended (cf. [45], paragraph 23), much like the range of intentions that underlie more general purposeful action. There is no finite list of discourse purposes, as there is, say, of syntactic categories. It remains an unresolved research question whether there is a finite description of the open-ended set of such intentions. However, even if there were finite descriptions, there would still be no finite list of intentions from which to choose. Thus, a theory of discourse structure cannot depend on choosing the DP/DSPs from a fixed list as is proposed in several alternative approaches [33, 36, 27], nor on the particulars of

[4]These two relations are similar to ones that play a role in parsing at the sentence level: immediate dominance and linear precedence. However, the dominance relation, like the one in Marcus and Hindle's D-theory [28], is partial (i.e., nonimmediate).

individual intentions. Although the particulars of individual intentions, like a wide range of common sense knowledge, are crucial to understanding any discourse, such particulars cannot serve as the basis for *determining* discourse structure.

What is essential for discourse structure is that such intentions bear certain kinds of structural relationships to one another. Since the CPs can never know the whole set of intentions that might serve as DP/DSPs, what they must recognize is the relevant structural relationships among intentions. Although there is an infinite number of intentions, there are only a small number of relations relevant to discourse structure that can hold between them.

2.3 Attentional State

The third component of discourse structure, the attentional state, is an abstraction of the participants' focus of attention as their discourse unfolds. The attentional state is a property of discourse, not of discourse participants. It is inherently dynamic, recording the objects, properties, and relations that are salient at each point in the discourse. The attentional state is modeled by a set of *focus spaces*; changes in attentional state are modeled by a set of transition rules that specify the conditions for adding and deleting spaces. We call the collection of focus spaces available at any one time the *focusing structure* and the process of manipulating spaces *focusing*.

The focusing process associates a focus space with each discourse segment; this space contains those entities that are salient--either because they have been mentioned explicitly in the segment or because they became salient in the process of producing or comprehending the utterances in the segment (as in Grosz' original work on focusing [16]). The focus space also includes the DSP; the inclusion of the purpose reflects the fact that the CPs are focused not only on what they are talking about but also on why they are talking about it.

To understand the attentional state component of discourse structure, it is important not to confuse it with two other concepts. First, the attentional state component is not equivalent to cognitive state, but is only one of its components. Cognitive state is a richer structure, one that includes at least the knowledge, beliefs,

desires, and intentions of an agent, as well as the cognitive correlates of attentional state as modelled in this paper. Second, although each focus space contains a DSP, the focus structure does *not* include the intentional structure as a whole.

Figure 1 illustrates how the focusing structure, in addition to modelling attentional state, serves during processing to coordinate the linguistic and intentional structures. The discourse segments (to the left of the figure) are tied to focus spaces (drawn vertically down the middle of the figure). The focusing structure is a stack. Information in lower spaces is usually accessible from higher ones (but less so than the information in the higher spaces); we will use a line with intersecting hash marks to denote when this is not the case. Subscripted terms are used to indicate the relevant contents of the spaces because representations of objects and not linguistic expressions are in the focus spaces.

Part one of Figure 1 shows the state of focusing when discourse segment DS2 is being processed. Segment DS1 gave rise to FS1 and had as its discourse purpose DSP_1. The properties, objects, relations, and purpose represented in FS1 are accessible but less salient than those in FS2. DS2 yields a focus space that is stacked relative to FS1 because DSP_1 in FS1 dominates DS2's DSP, DSP_2. As a result of the relationship between FS1 and FS2, reduced noun phrases will be interpreted differently in DS2 than in DS1. For example, if some red balls exist in the world one of which is represented in FS2 and another in FS1, then "the red ball" used in DS2 will be understood to mean that red ball that is represented in FS2. If, however, there is also a green truck (in the world) and it is represented only in FS1, "the green truck" occurring in DS2 will be understood as referring to that green truck.

Part two of Figure 1 shows the state of focusing when segment DS3 is processed. FS2 has been popped from the stack and FS3 has been pushed onto it because the DSP of FS3, DSP_3, is dominated solely by DSP_1, not by DSP_2. In this example, the dominance hierarchy includes only dominance relationships, though, in general, it may also include satisfaction-precedence relationships.

The stacking of focus spaces reflects the relative salience of the entities in each space during the corresponding segment's portion of the discourse. The stack relationships arise from the ways in which the various DSPs relate, information

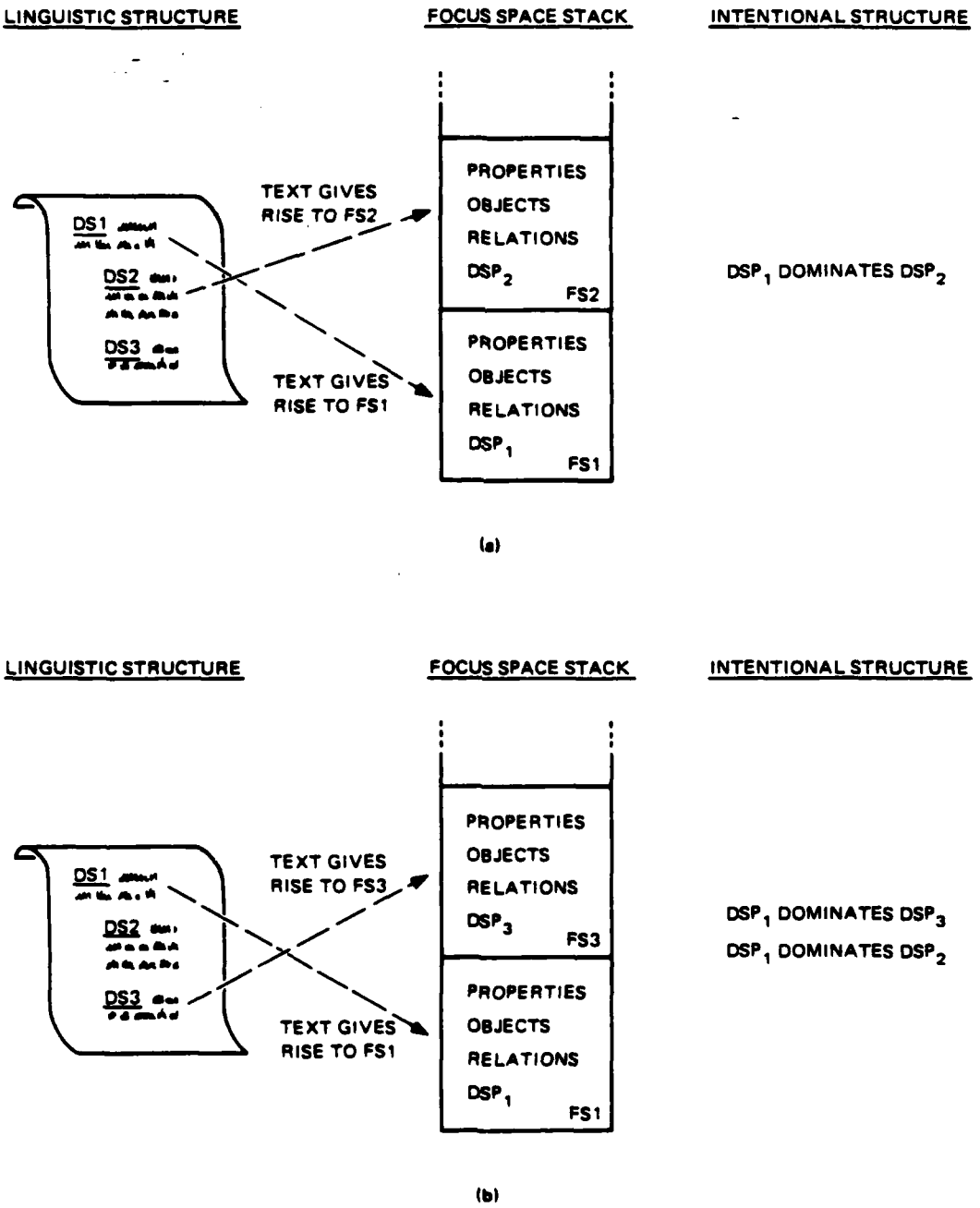


Figure 1: Discourse Segments, Focus Spaces and Dominance Hierarchy

represented in the dominance hierarchy (depicted on the right in the figure). The spaces in Figure 1 represent statically what results from a sequence of operations such as pushes onto and pops from a stack. A push occurs when the DSP for a new segment contributes to the DSP for the immediately preceding segment. When the DSP contributes to some intention higher in the dominance hierarchy, several focus spaces are popped from the stack before the new one is inserted.

Two essential properties of the focusing structure are now clear. First, the focusing structure is parasitic upon the intentional structure, in the sense that the relationships among DSPs determines pushes and pops. Note however, that the relevant operation may sometimes be indicated in the language itself. For example, the clue word "first" indicates the start of a segment whose DSP contributes to the DSP of the preceding segment. Second, the focusing structure, like the intentional and linguistic structures, evolves as the discourse proceeds. None of them exists a priori. Even in those rare cases in which an ICP has a complete plan for the discourse prior to uttering a single word, the intentional structure is constructed by the CPs as the discourse progresses. It may be more obvious that this is true for speakers and hearers of spoken discourse than *for* readers and writers of texts, but, even for the writer, the intentional structure is developed as the text is being written.

Figure 1 illustrates some fundamental distinctions between the intentional and attentional components of discourse structure. First, the dominance hierarchy provides, among other things, a complete record of the discourse-level intentions and their dominance (as well as, where relevant, satisfaction-precedence) relationships, whereas the focusing structure at any one time can contain essentially only information relevant to purposes in a portion of the dominance hierarchy. Second, at the conclusion of a discourse, if it completes normally, the focus stack will be empty, while the intentional structure will have been fully constructed. Third, when the discourse is being processed, only the attentional state can constrain the interpretation of referring expressions directly.

We can now also clarify some misinterpretations of focus-space diagrams and task structure in our earlier work [16, 18, 15]. The focus-space hierarchies in that work are best seen as representing attentional state. The task structure was used in two ways: (1) to represent common knowledge about the task; (2) as a special case of

the intentional structure we posit in this paper. Collapsing these roles was unfortunate as it fails to make an important distinction; furthermore, as is clear when moving to intentional structures more generally, it does not allow for differences between what one knows about a task and what intentions one has *for* (or makes explicit in discourse about) doing a task.

Although the representational scheme for encoding the focus–space hierarchies and the task structure was the same (partitioned networks [21]), the two structures were distinct. Several researchers (e.g., [24, 33]) misinterpreted the original research in an unfortunate and unintended way: they took the focus–space hierarchy to include (or be identical to) the task structure. The conflation of these two structures forces a single structure to contain information about attentional state, intentional relationships, and general task knowledge. It prevents a theory from accounting adequately for certain aspects of discourse, including interruptions (see Section 6).

A second confusion was to infer (incorrectly) that the task structure was necessarily a prebuilt tree. If the task structure is taken to be a special case of intentional structure, it becomes clear that the tree structure is simply a more constrained structure than one might require for other discourses; the nature of the task related to the task–oriented discourse is such that the intentional structure (i.e., dominance hierarchy) of the dialogue has both dominance and satisfaction–precedence relationships,[5] while other discourses may not be subject to significant precedence constraints among the DSPs. Furthermore, there has never been reason to assume that the task structures in task–oriented dialogues are prebuilt, any more than in the intentional structure of any other kind of discourse. It is rather that one objective of discourse theory (not a topic considered here, however) is to explain how the OCP builds up a model of the task structure by using information in the discourse.

In short, the focusing structure is the central repository for the contextual information needed to process utterances at each point in the discourse. It distinguishes those objects, properties, and relations most salient at that point and

---

[5]Even in the task case the orderings may be partial. In fact, the systems built for task–oriented dialogues [35, 44] did not use a prebuilt tree, but constructed the tree—based on a partially–ordered model—only as a particular discourse evolved.

has links to relevant parts of both the linguistic and intentional structures. During a discourse an increasing amount of information is discussed, only some of which continues to be needed for the interpretation of subsequent utterances. Hence, the ability to identify relevant discourse segments, the entitites they make salient, and their DSPs becomes more and more important. The role of attentional state in delineating the information necessary to understanding is thus central to discourse processing.

3 Two Examples

To illustrate the basic theory we have just sketched, we will give a brief analysis of two kinds of discourse: an argument from a rhetoric text and a task-oriented dialogue. For each example we will discuss the segmentation of the discourse, the intentions that underlie this segmentation, and the relationships among the various DSPs. In each case, we will point out some of the linguistic devices used to indicate segment boundaries as well as some of the expressions whose interpretations depend on those boundaries. The analysis is concerned with specifying certain aspects of the behavior to be explicated by a theory of discourse; the remainder of the paper provides a partial account of this behavior.

In the remainder of this paper we will distinguish between the *determination* of the DSP and the *recognition* of it. We will use the term *determination* to refer to a semantic-like notion, namely, the complete specification of what is intended by whom; we will use the term *recognition* to refer to a processing notion, namely, the processing that leads a discourse participant to identify what the intention is. These are obviously related concepts; the same information that determines a DSP may be used by an OCP to recognize it. However, some questions are relevant to only one of them. For example, the question of when the information becomes available is not relevant to determination but is crucial to recognition. An analogous distinction has been drawn with respect to sentence structure; the parse tree (determination) is differentiated from the parsing process (recognition) that produces the tree.

3.1 An Argument

Our first example is an argument taken from a rhetoric text [23];[6] it is an example used by Cohen [9] in her work on the structure of arguments. Figure 2 shows the dialogue and the eight discourse segments of which it is composed. The division of the argument into separate (numbered) clauses is Cohen's, but our analysis of the discourse structure is different. Although both analyses agree on the placement of utterance (4), some readers place this utterance in DS1 with utterances (1) through (3); this is an example of the kind of disagreement about boundary utterances found in Mann's data (as discussed in Section 2.1). The two placements lead to slightly different DSPs, but not radically different intentional structures. Because the differences do not affect the major thrust of the argument, we will discuss only one segmentation.

Figure 3 lists the primary component of the DSP for each of these segments and Figure 4 shows the dominance relationships that hold among these intentions. In Section 4 we discuss additional components of the discourse segment purpose; because these additional components are more important for completeness of the theory than for determining the essential dominance and satisfaction-precedence relationships between DSPs, we omit such details here. Rather than commit ourselves to a formal language in which to express the intentions of the discourse, we will use a shorthand notation and English sentences that are intended to be a gloss for a formal statement of the actual intentions.

All the primary intentions for this essay are intentions that the reader (OCP) come to believe some proposition. Some of these propositions, such as P5 and P6, can be read off the surface utterances directly. Other propositions and the intentions of which they are part, such as P2 and I2, are more indirect. Like the Gricean utterance-level intentions (the analogy with these will be explored in Section 4), DSPs may or may not be directly expressed in the discourse. In particular, they may be expressed in any of the following ways:

---

[6] The observant reader will note that this was written in the early days of the cinema, before the advent of sound; hence the quotation marks around ''movies.'' Note also that utterance (7) contains a somewhat odd preposition, and utterance (16) somewhat odd definite noun phrases. We have quoted the text exactly as it was printed.

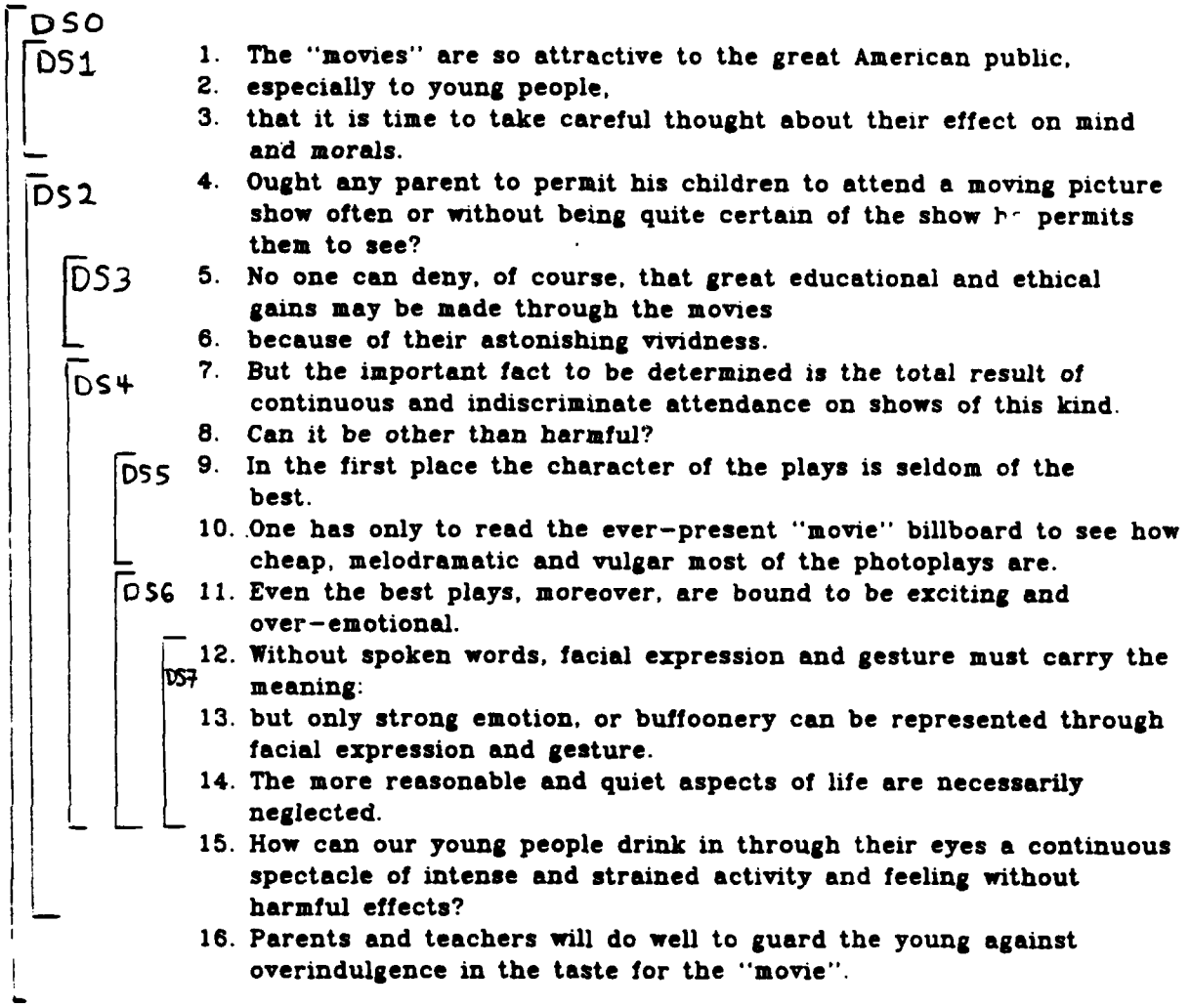


Figure 2: The Movies Essay

1. *explicitly* as in "I intend for you to believe that it's time to consider the effects of movies on mind and morals." [which would produce I1]

2. *directly, in one utterance*, as in (3) [which does produce I1]

3. *directly, through multiple utterances*, as in using (7) and the utterance "It can only be harmful" to produce I4.

4. *by derivation, in one or more utterances with an associated context*, as in (15) to produce I2.

Not only may information about the DSP be conveyed by a number of features of the utterances in a discourse, but it also may come in any utterance in a segment. For example, although I0 is the DP, it is stated directly only in the last utterance of

I0: (Intend S (Believe H P0))

where P0 = the proposition that parents and teachers should guard the young from overindulgence in the movies.

I1: (Intend S (Believe H P1))

where P1 = the proposition that it is time to consider the effect of movies on mind and morals.

I2: (Intend S (Believe H P2))

where P2 = the proposition that young people cannot drink in through their eyes a continuous spectacle of intense and strained activity without harmful effects.

I3: (Intend S (Believe H P3))

where P3 = the proposition that it is undeniable that great educational and ethical gains may be made through the movies.

I4: (Intend S (Believe H P4))

where P4 = the proposition that although there are gains, the total result of continuous and indisoriminate attendance at movies is harmful.

I5: (Intend S (Believe H P5))

where P5 = the proposition that the content of movies (i.e., the character of the plays) is not the best.

I6: (Intend S (Believe H P6))

where P6 = the proposition that the stories (i.e., the plays) in movies are exciting and over-emotional.

I7: (Intend S (Believe H P7))

where P7 = the proposition that movies portray strong emotion and buffoonery while neglecting the quiet and reasonable aspects of life.

Figure 3: Primary Intentions of the DSPs for Movies Essay

the essay. This leads to a number of questions about the ways in which OCPs can recognize discourse purposes, and about those junctures at which they need to do so. We turn to these matters directly in Section 5.

Dominance Relationships:

I0 DOM I1
I0 DOM I2
I2 DOM I3
I2 DOM I4
I4 DOM I5
I4 DOM I6
I6 DOM I7

Figure 4: Dominance Relationships for the DSPs of the Movies Essay

This discourse also provides several examples of the different kinds of interactions that can hold between the linguistic expressions in a discourse and the discourse structure. It includes examples of the devices that may be used to mark overtly the boundaries between discourse segments--examples of the use of aspect, mood, and particular "clue" words and phrases--as well as of the use of referring expressions that are affected by discourse segment boundaries.

The use of clue words and phrases to indicate discourse boundaries is illustrated in utterances (9) and (11); in (9) the phrase "in the first place" marks the beginning of DS5 while in (11) "moreover" ends DS5 and marks the start of DS6. These phrases also carry information about the intentional structure, namely, that DSP5 and DSP6 are dominated by DSP4, and that DSP5 satisfaction-precedes DSP6. In some cases, clue words and phrases have multiple functions; they convey propositional content as well as marking discourse segment boundaries. The "but" in utterance (7) is an example of such a multiple function use.

The boundaries between DS1 and DS2, DS4 and DS5, and DS4 and DS2 reflect changes of aspect and mood. The switch from declarative, present tense to interrogative modal aspect does not in itself seem to signal the boundary (for recognition purposes) in this discourse unambiguously, but it does indicate a possible line of demarcation which, in fact, is valid.

The effect of segmentation on referring expressions is shown by the use of the generic noun phrase "a moving picture show" in (4). Although a reference to the movies was made with a pronoun ("their") in (3), a full definite noun phrase is used in

(4). This use reflects and perhaps in part marks the boundary between the segments DS1 and DS2.

Finally, this discourse has an example of the trade-off between explicitly marking a discourse boundary, and recognizing it, as well as the relationship between the associated DSPs, by reasoning about the intentions themselves. There is no overt linguistic marker of the beginning of DS7; its separation must be inferred from DSP7 and its relationship to DSP6.

3.2 A Task-Oriented Dialogue

The second example is a fragment of a task-oriented dialogue taken from Grosz [18]; it is from the same corpus that was used by Grosz [15]. Figure 5 gives the dialogue fragment, and indicates the boundaries for its main segments.[7] Figure 6 gives the primary component of the DSPs for this fragment and shows the dominance relationships between them.

In contrast with the movies essay, the primary components of the DSPs in this dialogue are mostly intentions of the segment's ICP that the OCP intend to perform some action. Also unlike the essay, the dialogue has two agents initiating the different discourse segments. In this particular segment, the expert is the ICP of DS1 and DS5, while the apprentice is the ICP of DS2-4. To furnish a complete account of the intentional structure of this discourse, one must be able to say how the satisfaction of one agent's intentions can contribute to satisfying the intentions of another agent. Such an account is beyond the scope of this paper, but in Section 4 we discuss some of the complexities involved in providing one (as well as its role in discourse theory).

For the purposes of discussing this example, though, we need to stipulate two properties of the relationships among the participants' intentions. These properties seem rooted in features of cooperative behavior and depend on the two participants' sharing some particular knowledge of the task. First, it is a shared belief that, unless

---

[7] The segmentation omits some levels of detail. For example, utterances 19-24 are a segment within DS5. Rather than present this detail, we concentrate on the larger segments here so as to focus on the major issues with which this paper is concerned.

DS1
(1) E: First you have to remove the flywheel.
(2) A: How do I remove the flywheel?
(3) E: First, loosen the two allen head setscrews holding it to the shaft, then pull it off.
(4) A: OK.

DS2
(5) I can only find one screw. Where's the other one?
(6) E: On the hub of the flywheel.
(7) A: That's the one I found. Where's the other one?
(8) E: About ninety degrees around the hub from the first one.
(9) A: I don't understand. I can only find one. Oh wait, yes I think I was on the wrong wheel.
(10) E: Show me what you are doing.
(11) A: I was on the wrong wheel and I can find them both now.

DS3
(12) The tool I have is awkward. Is there another tool that I could use instead?
(13) E: Show me the tool you are using.
(14) A: OK.
(15) E: Are you sure you are using the right size key?
(16) A: I'll try some others.
(17) I found an angle I can get at it.

DS4
(18) The two screws are loose, but I'm having trouble getting the wheel off.

DS5
(19) E: Use the wheelpuller. Do you know how to use it?
(20) A: No.
(21) E: Do you know what it looks like?
(22) A: Yes.
(23) E: Show it to me please.
(24) A: OK.
(25) E: Good, Loosen the screw in the center and place the jaws around the hub of the wheel, then tighten the screw onto the center of the shaft. The wheel should slide off.

Figure 5: A Segment of a Task-Oriented Dialogue

he states otherwise, the OCP will adopt the intention to perform an action that the ICP intended him to. Second, in adopting the intention to carry out that action, the OCP also intends to do whatever subactions are necessary. Thus, once the apprentice intends to remove the flywheel, he also commits himself to intentions to loosen the setscrews and pull the wheel off. Note, however, that not all the subactions need to be introduced explicitly into the discourse. The apprentice may do several actions that are never mentioned, and the expert may assume these are undertaken on the basis of

Primary Intentions:

I1: (Intend E_{xpert} (Intend $A_{pprentice}$ (Remove A flywheel)))

I2: (Intend A (Intend E (Tell E A (Location other setscrew))))

I3: (Intend A (Intend E (Identify E A another tool)))

I4: (Intend A (Intend E (Tell E A (How (Getoff A wheel)))))

I5: (Intend E (Know-How-to A (Use A wheelpuller)))

Dominance Relationships:

I1 DOM I2
I1 DOM I3
I1 DOM I4
I4 DOM I5

Figure 6: Intentional Structure for the Task-Oriented Dialogue Segment

other information that the apprentice obtains. The partiality of the intentional structure stems in part from these characteristics of intentions and actions.

As in the movies essay, some of the DSPs for this dialogue are expressed directly in utterances. Utterances (1), (12), and (19) directly express the primary components of DSP1, DSP3, and DSP5, respectively. The primary component of DSP4 is a derived intention. The surface intention of "but I'm having trouble getting the wheel off" is that the apprentice intends the expert to believe that the apprentice is having trouble taking off the flywheel. I4 is derived from the utterance and its surface intention, as well as from features of discourse, conventions about what intentions are associated with the "I am having trouble doing X" type of utterance, and what the ICP and OCP know about the task that they have undertaken.

The dominance relationship that holds between I1 and I2, as well as the one that holds between I1 and I3, may seem problematic at first glance. It is not clear how locating any one setscrew contributes to removing the flywheel. It is even less clear

how, in and of itself, identifying another tool does. Two facts provide the link: first, that the apprentice (the OCP of DS1) has taken on the task of removing the flywheel; second, that the apprentice and expert share particular knowledge about the task. Note that some of this shared task knowledge comes from the discourse [e.g., utterance (3)], but some of it comes from general knowledge, perceptual information, and the like. Thus a combination of information is relevant to determining I2 and I3 and their relationships to I1, including all of the following: the fact that I1 is part of the intentional structure, the fact that the apprentice is currently working on satisfying I1, the utterance-level intentions of utterances (5) and (12), and general knowledge about the task.

Utterance (18) provides an example of the difference between the intentional structure and a general plan for the task. This utterance is part of DS4 and not just part of DS1 even though it contains references to more than one part of the overall task (which is what I1 is about). It functions to establish a new DSP, I4, as most salient. Rather than being regarded as a report on the overall status of the task, the first clause is best seen as modifying the DSP.[8] With it, the apprentice tells the expert that the trouble in removing the wheel is not with the screws. Thus, although general task knowledge is used in determining the intentional structure, it is not identical to it.

In this dialogue, there are fewer instances in which clue words are employed to indicate segment boundaries than in the movies essay. The primary example is the use of "first" in (1) to mark the start of the segment, and to indicate that its DSP is the first of several intentions whose satisfaction will contribute to satisfying the larger discourse of which it is a part.

The dialogue includes a clear example of the influence of discourse structure on referring expressions. The phrase "the screw in the center" is used in (25) to refer to the center screw of the wheelpuller, not one of the two setscrews mentioned in (18). This use of the phrase is possible because of the attentional state of the discourse structure at the time it is uttered.

---

[8]This "folding in" of an informing action to the request is similar to the action subsumption that Appelt [4] discusses in regard to referring expressions.

4 Some Properties of Discourse-Level Intentions

The intentions that serve as DP/DSPs are natural extensions of the intentions Grice [13] considers essential to developing a theory of utterer's meaning. However, there is a crucial difference between our use of discourse-level intentions in this paper (and the theory as developed so far) and Grice's use of utterance-level intentions. We will not address the issue of discourse **meaning**,[9] but will focus on the role of DP/DSPs in determining discourse **structure** and on specifying how these intentions can be recognized by an OCP. Although the intentional structure of a discourse plays a role in determining discourse meaning, the DP/DSPs are not in and of themselves discourse segment meaning. The connection between intentional structure and discourse meaning is similar to that between attentional and cognitive states; the attentional state plays a role in a hearer's understanding what the speaker means by a particular sequence of utterances in a discourse segment, but it is not the only aspect of cognitive state that contributes to this understanding.

We will draw on some particulars of Grice's definition of utterer's meaning to explain DSPs more fully. His initial definition is as follows:

> "*U* meant something by uttering *x* is true iff [for some audience *A*]:
>
> 1. *U* intended, by uttering *x*, to induce a certain response in *A*
> 2. *U* intended *A* to recognize, *at least in part from the utterance of x*, that *U* intended to produce that response
> 3. *U* intended the fulfillment of the intention mentioned in (2) to be at least in part *A*'s reason for fulfilling the intention mentioned in (1)."

Grice refines this definition to address a number of counterexamples. The

---

[9] This is not to say we think discourse meaning to be either an unimportant or a solved problem. Quite the contrary. However, an adequate theory of discourse meaning will need to rest in part on an adequate theory of discourse structure. Our current concern is with this latter problem.

following portion of his final definition[10] is relevant to this paper:

"By uttering x U meant that $^*\psi p$ is true iff

($\exists A$)($\exists f$ [features of the utterance]) ($\exists c$ [ways of correlating f with utterances[11]]):

(a) U uttered x intending

1. A to think x possesses f
2. A to think f correlated in way c with ψ-*ing* that p
3. A to think, on the basis of fulfillment of (1) and (2) that U intends A to think that U ψ's that p
4. A on the basis of fulfillment of (3) to think that U ψ's that p
5. and (in some cases), A on the basis of fulfillment of (4) himself to ψ that p."

Grice takes $^*\psi p$ to be the meaning of the utterance where $^*\psi$ is a mood indicator associated with the propositional attitude ψ (e.g., $^*\psi$=assert and ψ=believe). He considers attitudes like believing that S is a German soldier and intending to give ICP a beer as examples of the kinds of ψ-*ing* that p that utterance intentions can embed. For expository purposes we will use the following notation to represent these utterance-level intentions:

Intend(ICP, Believe(OCP, ICP is a German soldier))
Intend(ICP,Intend(OCP, OCP give ICP a beer))

To extend Grice's definition to discourses, we replace the utterance x with a discourse segment DS, the utterer U with the initiator of a discourse segment ICP, and the audience A with the OCP. To complete this extension, a number of issues must be addressed including the following: (1) specifying the discourse-level intentions and

---

[10] We are using Redefinition IVB: a further redefinition deals with abstracting about audience and would unnecessarily complicate our initial picture of intentions and discourse.

[11] Grice [13] mentions iconic, conventional, and associative modes, giving examples of each.

attitudes that correspond to the utterance-level intentions and ψ's *that p*; (2) identifying the kinds of *f*s that are correlated with these intentions; (3) determining the ways in which the *f*s and the intentions are correlated; (4) specifying how the discourse-level intentions can be recognized by an OCP. The proper treatment of (4) is especially necessary for a computationally useful account of discourse. Each of these issues corresponds to an unresolved problem in discourse theory. In this paper we can only impose some constraints on the solutions and make some suggestions as to which approaches might be tried.

At the discourse level, just as at the utterance level, the intended recognition of intentions plays a central role. It is important to distinguish effects that are intended to be recognized from other intended effects that do not need to be recognized. For example, a compliment achieves its intended effect only if the intention to compliment is recognized; in contrast, a scream of "boo" typically achieves its intended effect (scaring the hearer) without the hearer having to recognize the speaker's intention. The DSPs are intended to be recognized: they achieve their effects, in part, because the OCP recognizes the ICP's intention for the OCP to ψ *that p*. The OCP's recognition of this intention is crucial to its achieving the desired effect.

4.1 The Basic Generalization

In extending Grice's analysis to the discourse level, we have to consider not only individual beliefs and intentions, but also the relationships among them that arise because of the relationships among various discourse segments (and utterances within a segment) and the purposes the segments serve with respect to the entire discourse. To clarify these relationships, consider an analogous situation with nonlinguistic actions.[12] An action may divide into several subactions; for example, the planting of a rose bush divides into preparing the soil, digging a hole, placing the rose bush in the hole, filling the rest of the hole with soil, and watering the ground around the bush.

---

[12]This analogy is meant to help clarify and motivate the discussion. Although it also suggests some important problems in common between research on discourse and research on theories of action and intention, those issues are the subject of another paper. Section 9 discusses some of the still unresolved problems we take to be most crucial.

The intention to perform the planting action includes several subsidiary intentions (one for each of the subactions--namely, to do it).

In discourse, in a manner that is analogous to nonlinguistic actions, the DP (and some DSPs) includes several subsidiary intentions related to the DSPs it dominates. For purposes of exposition, we will use the term *primary intention* to distinguish the overall intention of the DP from the subsidiary intentions of the DP. For example in the movies argument of Section 3.1, the primary intention is for the reader to come to believe that parents and teachers should keep children from seeing too many movies; in the task dialogue of Section 3.2, the intention is that the apprentice remove the flywheel. Subsidiary intentions include, respectively, the intention that the reader believe that it important to evaluate movies and the intention that the expert help the apprentice locate the second setscrew.

The discourse situation differs from the general-action situation in two important ways. In a discourse, the ICP also intends the OCP to recognize the ICP's beliefs about the connections among various propositions and actions. For example, in the movies argument, the reader (OCP) is intended to recognize that the author (ICP) believes some propositions provide support for others; in the task dialogue the expert (ICP) intends the apprentice (OCP) to recognize that the expert believes the performance of certain actions contributes to the performance of other actions. In contrast, in the general-action situation in which there is no communication, there is no need for recognition of another agent's beliefs about the interrelationship of various actions and intentions.

The discourse situation differs from the analogous general action situation in a second way. To perform some action, the agent (or, in some cases, agents) must perform each of the subactions involved; by performing all of these subactions the agent performs the action. In contrast in a discourse, the participants share the assumption of *discourse sufficiency*: it is a convention of the communicative situation that the ICP believes the discourse is sufficient to achieve the primary intention of the DP. Discourse sufficiency does not entail logical sufficiency or action completeness. It is not necessarily the case that satisfaction of all of the DSPs is sufficient in and of itself for satisfaction of the DP. Rather, there is an assumption that the information conveyed in the discourse will suffice *in conjunction with other*

information the ICP believes the OCP has (or can obtain) to allow for satisfaction of the primary intention of the DP. Satisfaction of all of the DSPs, in conjunction with this additional information, is enough for satisfaction of the DP. Hence, in discourse the intentional structure (the analogue of the action hierarchy) need not be complete.

For example, the propositions expressed in the movies essay do not provide a logically sufficient proof of the claim. The author furnishes information that he believes to be adequate for the reader to reach the desired conclusion and assumes the reader will supplement what is actually said with appropriate additional information and reasoning. Likewise, the task dialogue does not mention all the subtasks explicitly. Instead, the expert and apprentice discuss explicitly only those subtasks for which some instruction is needed or in connection with which some problem arises.

To be more concrete, we will look at the extension of the Gricean analysis for two particular cases, one involving a belief, the other an intention to perform some action. We will consider only the simplest situations, in which the primary intentions of the DP/DSPs are about either beliefs or actions, but not a mixture. Although the task dialogue obviously involves a mixture, this is an extremely complicated issue that demands additional research.

4.2 The Belief Case

In the belief case, the primary intention of the DP is to get the OCP to believe some proposition, say p. Each of the discourse segments is also intended to get the OCP to believe a proposition, say q_i for some i=1,...,n (where there are n discourse segments). In addition to the primary intention--i.e., that the OCP should come to believe p--the DP includes an intention that the OCP come to believe each of the q_i and, in addition, an intention that the OCP come to believe the q_i provide support for p. We can represent this schematically as:[13],

$\forall i=1,...,n$ Intend(ICP, Believe(OCP,p) $\wedge$

---

[13]Here again we use a notational shorthand rather than a formal language to make some of the relationships clearer.

Believe(OCP,q_i) $\wedge$
Believe(OCP, Supports (p, $q_1 \wedge \ldots \wedge q_n$)))

There are several things to note here. To begin with, the first intention, (Intend ICP (Believe (OCP p)), is the primary component of the DSP. Second, each of the intended beliefs in the second conjunct corresponds to the primary component of some embedded DSP. Third, the *supports* relation is not implication. The OCP is not intended to believe that the q_i imply p, but rather to believe that the q_i in conjunction with other facts and rules that the ICP assumes the OCP has available or can obtain and thus come to believe are sufficient for the OCP to conclude p. Fourth, the DP/DSP may only be completely determined at the end of the discourse (segment); we discuss the effect of this on recognition in the next section.

Finally, to determine how the discourse segments corresponding to the q_i are related to the one corresponding to p, the OCP only has to believe that the ICP believes a supports relationship holds. Hence, for the purpose of recognizing the discourse structure, it would be sufficient for the third clause to be

. . . Believe(OCP, Believe(ICP, Supports (p, $q_1 \wedge \ldots \wedge q_n$)))

However, the DP of a belief-case discourse is not merely to get the OCP to believe p, but to get the OCP to believe p by virtue of believing the q_i. That this is so can be seen clearly by considering situations in which the OCP already believes p and is known by the ICP to do so, but does not have a good reason for believing p. This last property of the belief case is not shared by the action case.

There is an important relationship between the supports relation and the dominance relation that can hold between DP/DSPs; it is captured in the following rule (using the same notation as above):

$\forall i=1,\ldots,n$ Intend(CP_1, Believe(CP_2,p)) $\wedge$
Intend(CP_1, Believe(CP_2,q_i)) $\wedge$
Believe(CP_1, Supports(p, $q_1 \wedge \ldots \wedge q_n$)) $\leftrightarrow$
DOM(Intend(CP_1, Believe(CP_2,p))
Intend(CP_1, Believe(CP_2,q_i)))

The implication in the forward direction states that if a conversational participant (CP_1) believes that the proposition p is supported by the proposition q_i, and he intends another participant (CP_2) to adopt these beliefs, then his intention

that CP_2 believe p dominates his intention that CP_2 believe q_i. Viewed intuitively, CP_1's belief that q_i provides support for p, underlies his intention to get CP_2 to believe p by getting him to believe q_i. The satisfaction of CP_1's intention that CP_2 should believe q_i will help satisfy CP_1's intention that CP_2 believe p. This relationship plays a role in the recognition of DSPs.

4.3 The Action Case

An analogous situation holds for a discourse segment comprising utterances intended to get the OCP to perform some set of actions directed at achieving some overall task (e.g., some segments in the task-oriented dialogue of Section 3.2). The full specification of the DP/DSP contains a *generates* relation that is derived from a relation defined by Goldman [12]. For this case, the DP/DSPs are of the following form:

$$\begin{aligned}\forall i=1,\ldots,n\ \text{Intend}(&\text{ICP, Intend}(\text{OCP, Do}(A)) \wedge \\ &\text{Intend}(\text{OCP, Do}(a_i)) \wedge \\ &\text{Believe}(\text{OCP, Believe}(\text{ICP, Generates}(A, a_1 \wedge \ldots \wedge a_n))))\end{aligned}$$

Each intention to act represented in the second conjunct corresponds to the primary intention of some discourse segment.

Like supports, the generates relation is partial (its partiality distinguishes it in part from Goldman's relation). Thus, the OCP is not intended to believe that the ICP believes that performance of a_i alone is sufficient for performance of A, but rather that doing all of the a_i and other actions that the OCP can be expected to know or figure out constitutes a performance of A. In the task dialogue of Section 3.2 many actions that are essential to the task (e.g., the apprentice picking up the Allen wrench and applying it correctly to the setscrews) are never even mentioned in the dialogue.

Note that it is unnecessary for the ICP or OCP to have a complete plan relating all of the a_i to A at the start of the discourse (or discourse segment). All that is required is that, for any given segment, the OCP be able to determine what intention to act the segment corresponds to and which other intentions dominate that intention. Finally, unlike the belief case, the third conjunct here requires only that the OCP recognize that the ICP believes a generates relationship holds. The OCP can do A by

virtue of doing the a_i without coming himself to believe anything about the relationships between A and the a_i.

As in the belief case, there is an equivalence that links the generates relation among actions to the dominance relation between intentions. Schematically, it is as follows:

$$\begin{aligned}\forall i=1,\ldots,n\ [&\text{Intend}(CP_1,\ \text{Intend}(CP_2,\ \text{Do}(A))) \wedge \\ &\text{Intend}(CP_1,\ \text{Intend}(CP_2,\ \text{Do}(a_i))) \wedge \\ &\text{Believe}(CP_1,\ \text{Generates}(A,\ a_1 \wedge \ldots \wedge a_n))] \leftrightarrow \\ &\text{DOM}(\text{Intend}(CP_1,\ \text{Intend}(CP_2,\ \text{Do}(A))) \\ &\quad \text{Intend}(CP_1,\ \text{Intend}(CP_2,\ \text{Do}(a_i))))\end{aligned}$$

This equivalence states that, if an agent (CP_1) believes that the performance of some action (a_i) contributes in part to the performance of another action (A), and if CP_1 intends for CP_2 to (intend to) do both of these actions, then his intention that CP_2 (intend to) perform a_i is dominated by his intention that CP_2 (intend to) perform A. Viewed intuitively, CP_1's belief that doing a_i will contribute to doing A underlies his intention to get CP_2 to do A by getting CP_2 to do a_i. The satisfaction of CP_1's intention for CP_2 to do a_i will help satisfy CP 's intention for CP_2 to do A.

So, for example, in the task-oriented dialogue of Section 3.2, the expert knows that using the wheelpuller is a necessary part of removing the flywheel. His intention that the apprentice intend to use the wheelpuller is thus dominated by his intention that the apprentice intend to take off the flywheel. Satisfaction of the intention to use the wheelpuller will contribute to satisfying the intention to remove the flywheel. In general, the action a_i does not have to be a necessary action though it is in this example (at least if the task is done correctly).

A definitive statement characterizing primary and subsidiary intentions for task-oriented dialogues awaits further research not only in discourse theory, but also in the theory of intentions and actions. In particular, a clearer statement of the interactions among the intentions of the various discourse participants (with respect to both linguistic and nonlinguistic actions) awaits the formulation of a better theory of cooperation and multiagent activity.

5 Recognition Issues

In this section we consider a number of issues related to the question of how a discourse participant (in particular the OCP) recognizes the DSP of a given segment. In previous sections of the paper we abstracted from the cognitive states of the discourse participants. The various components of discourse structure discussed so far are properties of the discourse itself, not of the discourse participants. The validity of this theory as a computational theory (as well as its potential usefulness for natural-language processing) rests ultimately on whether one can specify the role of the different components in discourse processing. In particular, it is necessary to provide computationally tractable algorithms for recognizing DP/DSPs and for constructing computational models of attentional state and intentional structure for use in understanding and producing discourse.

Viewed from this perspective, the most problematic structure for recognition is the intentional one. If, as we have claimed, for the discourse to be coherent and comprehensible, the OCP must be able to recognize both the DP/DSPs and relationships (dominate, satisfaction-precede) between them, then the question of how the OCP does so is a central issue. Although we will not be able to answer this question completely, we can delimit particular issues to be addressed and provide certain constraints on the problem. The discussion will address three closely related issues: what specifically the OCP must recognize, what information the OCP can utilize in effecting the recognition, and when that information becomes available.

5.1 What Must Be Recognized

For the discourse as a whole, as well as each of its segments, the OCP must identify the intention that serves as the discourse segment purpose[14] and its

---

[14]We assume here that the OCP must recognize intentions rather than actions. The argument that such is the case is beyond the scope of this paper. At a very general level, it centers on the possibility of the exact same sequence of [utterance] actions corresponding to two different discourse structures where the difference is statable only in terms of the ICP's intentions. The possibility of such sequences was suggested to us by Michael Bratman [personal communication].

relationship to other discourse-level intentions. In particular, the OCP must be able to recognize what other DSPs that specific intention dominates and is dominated by, and, where relevant, to which other DSPs it bears satisfaction-precedence relationships. Thus, there are two closely associated recognition problems: what the intention is, and what intentions it is related to.

One role of the attentional state is to constrain the range of DSPs considered as candidates for domination or satisfaction-precedence of the DSP of the current segment. Only those DSPs in some space on the focusing stack are viable prospects. As a result of this use of the focusing structure, the theory predicts that this decision will be a local one with respect to attentional state. Because two focus spaces may be close to each other in the attentional structure without the discourse segments they arise from necessarily being close to one another and vice versa, this prediction corresponds to a claim that locality in the focusing structure is what matters to determination of the intentional structure.

A number of alternative theories (e.g., [14, 22, 27, 33]) claim that a set of rhetorical relations underlie discourse structure. The different theories identify different specific relations, but they all use them in a recognition role similar to the ones played by DP/DSPs and the dominance and satisfaction-precedence relationships between them in this theory. Among the various rhetorical relations that have been investigated are elaboration, summarization, enablement, justification, and challenge. The intentions that typically serve as DP/DSPs are more basic than those that underlie such rhetorical relations. They are also not specialized for linguistic behavior; their satisfaction can be realized by nonlinguistic actions as well as linguistic ones.

Rhetorical relationships do not have a privileged role in the account given here for several reasons. Although they appear to provide a metalevel description of the discourse, their role in discourse interpretation remains unclear. With respect to discourse processing, it seems obvious that the ICP and OCP have very different access to them. In particular, the ICP may well have such rhetorical relationships "in mind" as he produces utterances, whereas it is much less clear when (if at all) the OCP infers them. A claim of the theory being developed in this paper is that a discourse can be understood at a basic level even if the OCP never does or is able to construct,

let alone name, such rhetorical relationships. Furthermore, we conjecture that these relationships could be recast as a combination of domain specific information and DSP relationships (DOM and SP). Even so, rhetorical relationships are quite likely useful to the theoretician as an analytical tool for some aspects of discourse analysis.[15]

5.2 Information Constraining the DSP

At least three different kinds of information play a role in the determination of the DSP: specific linguistic markers, utterance-level intentions, and general knowledge about actions and objects in the domain of discourse. These are among the f's [features] of the Gricean analysis. Each plays a part in the OCP's recognition of the DSP and can be utilized by the ICP to facilitate this recognition.

The most distinguished linguistic means that speakers have for indicating discourse segment boundaries and conveying information about the DSP are the clue words and clue phrases described by Cohen [9], Grosz [16], Reichman [32], and Polanyi and Scha [29].[16] Because some clue words may be used as clausal connectors, there is a need to distinguish their discourse use from their use in conveying propositional content at the utterance level. For example, the word "but" functions as a boundary marker in utterance (7) of the discourse in Section 3.1, but it can also be used solely (as in the current utterance) to convey propositional content (e.g., the conjunction of two propositions) and serve to connect two clauses within a segment.

As discussed in Section 7, clue phrases can provide information about dominance and satisfaction-precedence relationships between segments' DSPs. However, they may not completely specify which DSP dominates or satisfaction-precedes the DSP of the segment they start. Furthermore, clue phrases that explicitly convey information only about the attentional structure (see Section 7) may be ambiguous about the state to

---

[15]This claim reflects a move analogous to the one made by Cohen and Levesque [10] in showing that the definitions of various speech acts can be derived as lemmas within a general theory of rational behavior.

[16]Additional explicit markers, linguistic and extralinguistic, that have been studied to a lesser extent include tense and aspect, intonation, tone of voice, gesture, and eye gaze.

which attention is to shift. For example, if there have been several interruptions (see Section 6), the phrase "but anyway" indicates a return to some previously interrupted discourse, but does not specify which one. Although clue words and phrases do not completely specify a DSP, the information they provide is useful in limiting the options to be considered.

The second kind of information the OCP has available is the utterance-level intention of each utterance in the discourse. As the discussion of the movies example (Section 3.1) pointed out, the DSP may be identical to the utterance-level intention of some utterance in the segment. Alternatively, the DSP may combine the intentions of several utterances, as is illustrated in the following discourse segment:

I want you to arrange a trip for me to Palo Alto.
It will be for two weeks.
I only fly on TWA.

The DSP for this segment is, roughly, that the ICP intends for the OCP to make (complete) trip arrangments for the ICP to go to Palo Alto for two weeks, under the constraint that any flights be on TWA. The Gricean intentions for these three utterances are as follows:

Utterance1: ICP intends that OCP believe that ICP intends that
OCP intend to make trip plans for ICP to go
to Palo Alto

Utterance2: ICP intends that OCP believe that ICP intends
OCP to believe that the trip will last two weeks

Utterance3: ICP intends that OCP believe that ICP intends
OCP to believe that ICP flies only on TWA

These intentions must be combined in some way to produce the DSP. The process is quite complex, since the OCP must recognize that the reason for utterances 2 and 3 is not simply to have some new beliefs about the ICP, but to use those beliefs in arranging the trip. While this example fits the schema of a request followed by two informings, schemata will not suffice to represent the behavior as a general rule. A different sequence of utterances with different utterance-level intentions can have the same DSP; this is the case in the following segment:

S1: Have I told you yet to arrange my trip to
Palo Alto? Remember that I will fly

only on TWA. OK?
S2: OK.
S3: I'm planning on staying for two weeks.

It is possible for a sequence that consists of a request followed by two informings not to result in a modification of the trip plans. For example, in the following sequence the third utterance results in changing the way the arrangements are made, rather than constraining the nature of the arrangements themselves.

I want you to arrange a two-week trip for me to Palo Alto.
I fly only on TWA.
The rates go up tomorrow, so you'll want to call today.

Not only is the contribution of utterance-level intentions to DSPs complicated, but in some instances the DSP for a segment may both constrain and be partially determined by the Gricean intention for some utterance in the segment. For example, the Gricean-intention for utterance (15) in the movies example (Section 3.1) is derived from a combination of facts about the utterance itself, and from its place in the discourse. On the surface, (15) appears to be a question addressed to the OCP; its intention would be roughly that the ICP intends the OCP to believe that the ICP wants to know how young people, etc. But (15) is actually a rhetorical question and has a very different intention associated with it--namely, that the ICP intends the OCP to believe proposition P2 (namely, that young people cannot drink in through their eyes a continuous spectacle of intense and strained activity without harmful effects). In this example, this particular intention is also the primary component of the DSP.

The third kind of information that plays a role in determining the DP/DSPs is shared knowledge about actions and objects in the domain of discourse. This shared knowledge is especially important when the linguistic markers and utterance-level intentions are insufficient for determining the DSP precisely.

In Section 4 we presented two rules stating equivalences; one linked a dominance relation between two DSPs with a supports relation between propositions and the other linked a dominance relation between DSPs to a generates relation between actions. Use of these rules in one direction allows for (partially) determining what supports or generates relationship holds from the dominance relationship. But the rules can be used in the opposite direction also: if, from the content of utterances and reasoning

about the domain of discourse, a supports or generates relationship can be determined, then the dominates relationship between DSPs can be determined. In such cases it is important to derive the dominance relationship so that the appropriate intentional and attentional structures are available for processing or determining the interpretation of the subsequent discourse.

From the perspective of recognition, a tradeoff implicit in the two equivalences is important. If the ICP makes the dominance relationship between two DSPs explicit (e.g., with clue words), then the OCP can use this information to help recognize the (ICP's beliefs about the) supports relationship. Conversely, if the ICP's utterances make clear the (ICP's beliefs about the) supports or dominance relationship, then the OCP can use this information to help recognize the dominance relationship. Although it is most helpful to use the dominance relationships to constrain the search for appropriate supports and generates relationships, sometimes these relationships can be inferred reasonably directly from the utterances in a segment using general knowledge about the objects and actions in the domain of discourse. It remains an open question what inferences are needed and how complex it will be to compute supports and generates relationships if the dominance relationship is not directly indicated in a discourse.

Utterances from the movies essay illustrate this tradeoff. In utterance (9), the phrase "in the first place" expresses the dominance relationship between DSPs of the new segment DS5 and the parent segment DS4 directly. Because of the dominance relationship (as well as the intentions expressed in the utterances), the OCP can determine that the ICP believes that the proposition that the content of the plays is not the best provides support for the proposition that the result of indiscriminate movie going is harmful. Hence determining dominance yields the support relation. The support relation can also yield dominance. Utterances (12)–(14), which comprise DS7, are not explicitly marked for a dominance relation. It can be inferred from the fact that the propositions in (12)–(14) provide support for the proposition embedded in DSP6 (that is, that the stories in movies are exciting and over-emotional) that DSP6 dominates DSP7.

The inference of relationships like supports and generates is simpler than that of rhetorical relationships proposed in other theories (e.g., elaboration, justification)

in two ways. First, the supports and generates relations themselves are simpler and more basic. Second, supports and generates relationships hold between propositions or actions in the domain of discourse; they depend on facts of how the world is, not on facts of the discourse. In contrast, the rhetorical relations combine discourse and domain information.

Finally, we conjecture that the more information an ICP supplies explicitly in the actual utterances of a discourse, the less reasoning an OCP has to do to achieve recognition. Cohen [9] has ventured the same hypothesis regarding the problem of recognizing the relationship between one proposition and another. We believe these conjectures are related. Several kinds of information typically provide partial information about these relationships; they are each partially constraining, but only in their ensemble do they constrain in full. To the extent that more information is furnished by one, less is needed from the others. It is the combination of linguistic markers of discourse structure, the particular discourse-level intentions currently in focus, and the shared knowledge of certain portions of the discourse content that enables the OCP to recognize DSP.

5.3 When Is the Intention Recognized?

As discussed in Section 2.2, the intentional structure evolves as the discourse does. By the same token, the discourse participants' mental-state correlates of the intentional structure are not prebuilt; neither participant may have a complete model of the intentional structure "in mind" until the discourse is completed. The dominance relationships that actually shape the intentional structure cannot be known a priori, because the specific intentions that will come into play are not known (never by the OCP, hardly ever by the ICP) until the utterances in the discourse have been made. Although it is assumed that the participants' common knowledge includes enough information about the domain to determine various relationships such as supports and generates, it is not assumed that, prior to a discourse, they actually had inferred and are aware of all the relationships they will need for that discourse.

Because any of the utterances in a segment may contribute information relevant to a complete determination of the DSP, the recognition process is not complete until the end of the segment. However, the OCP must be able to recognize at least an

abstraction of the DSP so that he can make the proper moves with respect to the attentional structure. That is, some combination of explicit indicators and intentional and propositional content must allow the OCP to ascertain where the DSP will fit in the intentional structure at the beginning of a segment, even if the specific intention that is the DSP cannot be determined until the end of the segment.

Utterance (15) in the movies example illustrates this point. The author writes, "How can our young people drink in through their eyes a continuous spectacle of intense and strained activity and feeling without harmful effects?" The primary intention I2 is derived from this utterance, but this cannot be done until very late in the discourse segment [since (15) occurs at the end of DS2]. Furthermore, the segment for which I2 is primary has complex embedding of other segments. Utterance (16), intention I0, and DS0 constitute another example of the expression of a primary intention late in a discourse segment. In that case, I0 cannot be computed until (16) has been read, and (16) is not only the last utterance in DS0, but is one that covers the entire essay. If an OCP must recognize a DSP to understand a segment, then we ask: how does the OCP recognize a DSP when the utterance from which its primary intention is derived comes so late in the segment?

We conjecture with regard to such segments as D2 of the movies essay that the primary intention (e.g., I2) may be determined approximately (and hence a generalized version become recognizable) before the point at which it is actually expressed in the discourse. While the DP/DSP may not be expressed early, there is still *partial information* about it. This partial information often suffices to establish dominance (or satisfaction-precedence) relationships for additional segments. As these latter are placed in the hierarchy, their DSPs can provide further partial information for the underspecified DSP. For example, even though the intention I0 is expressed directly only in the last utterance of the movies essay, utterance (4) expresses an intention to know whether p or ~p is true (i.e., whether or not parents should let children see movies often and without close monitoring). I0 is an intention to believe, whose proposition is a generalization of the ~p expressed in (4). Consider also the primary intention I4. It occurs in a segment embedded within DS2, is more general than I2, but is an approximation to it. It would not be surprising to discover that OCPs can in fact predict something close to I2 on the basis of I4, utterances (9)-(14), and the partial dominance hierarchy available at each point in the discourse.

6 Application of the Theory: Interruptions

Interruptions in discourses pose an important test of any theory of discourse structure. Because processing an utterance requires ascertaining out how it fits with previous discourse, it is crucial to decide which parts of the previous discourse are relevant to it, and which cannot be. Interruptions, by definition, do not fit; consequently their treatment has implications for the treatment of the normal flow of discourse. Interruptions may take many forms--some are not at all relevant to the content and flow of the interrupted discourse, others are quite relevant, and many fall somewhere inbetween these extremes. A theory must differentiate these cases and explain (among other things) what connections exist between the main discourse and the interruption, and how the relationship between them affects the processing of the utterances in both.

The importance of distinguishing between intentional structure and attentional state is evident in the three examples considered in Sections 6.2, 6.3, and 6.4. The distinction also permits us to explain a type of behavior deemed by others to be similar--so-called semantic returns--an issue we examine in Subsection 6.5.

These examples do not exhaust the types of interruptions that can occur in discourse. There are other ways to vary the explicit linguistic (and nonlinguistic) indicators used to indicate boundaries, the relationships between DSPs, and the combinations of focus space relationships present. However, the examples provide illustrations of interruptions at different points along the spectrum of relevancy to the main discourse. Because they can be explained more adequately by the theory of discourse structure presented here than by previous theories, they support the importance of the distinctions we have drawn.

6.1 Preliminary Definitions

From an informal view, we observe that interruptions are pieces of discourse that break the flow of the preceding discourse. An interruption is in some way distinct from the rest of the preceding discourse; after the break for the interruption, the discourse returns to the interrupted piece of discourse. In the example below,

from Polanyi and Scha [31], there are two (separate) discourses, D1 indicated in normal type, and D2 in italics. D2 is an interruption that breaks the flow of D1 and is distinct from D1.

D1: John came by
and left the groceries
D2: *Stop that*
you kids
D1: and I put them away
after he left

Using the theory described in previous sections, we can capture the above intuitions about the nature of interruptions with two slightly different definitions. The strong definition holds for those interruptions we classify as "true interruptions," but the weaker form is needed for other types. The two definitions are as follows:

Strong definition: An interruption is a discourse segment whose DSP is not dominated by the DSP of any preceding segment.

Weak definition: An interruption is a discourse segment whose DSP is not dominated by the DSP of the immediately preceding segment.

Neither of the above definitions includes an explicit mention of our intuition that there is a "return" to the interrupted discourse after an interruption. The return is an effect of the normal progress of a conversation. If we assume a focus space is normally popped from the focus stack if and only if a speaker has satisfied the DSP of its corresponding segment, then it naturally follows both that the focus space for the interruption will be popped *after* the interruption and that the focus space for the interrupted segment will be at the top of the stack because its DSP is yet to be satisfied.

There are other kinds of discourse segments that one may want to consider in light of the interruption continuum and these definitions. Clarification dialogues [1] and debugging explanations [42] are two such possibilities. Both of them, unlike the interruptions discussed here, share a DSP with their preceding segment and thus do not conform to our definition of interruption. We conjecture that these kinds of discourses constitute another general class of discourse segments that, like interruptions, can be abstractly defined.

6.2 Type 1: True Interruptions

The first kind of interruption is the true interruption, which follows the strong definition of interruptions. It is exemplified by the interruption given in the previous subsection. Discourses D1 and D2 have distinct, unrelated purposes and convey different information about properties, objects, and relations. Since D2 is embedded within D1, one expects the discourse structures for the two segments to be somehow embedded as well. The theory described in this paper differs from Polanyi and Scha's [30] (and other more radically different proposals as well; e.g., [24, 9, 33]) because the embedding occurs *only* in the attentional structure. As shown in Figure 7, the focus space for D2 is pushed onto the stack above the focus space for D1, so that the focus space for D2 is more salient than the one for D1, until D2 is completed. The intentional structures for the two segments are distinct. There are two DP/DSP structures for the utterances in this sequence--one for those in D1 and the other for those in D2. It is not necessary to relate these two; indeed, from an intuitive point of view, they are not related.

The focusing structure for true interruptions is different from that for the normal embedding of segments, because the focusing boundary between the interrupted discourse and the interruption is impenetrable.[17]. (This is depicted in the figure by a line with intersecting hash marks between focus spaces). The impenetrable boundary between the focus spaces prevents entities in the spaces below the boundary from being available to the spaces above it. Because the second discourse shifts attention totally to a new purpose (and may also shift the identity of the intended hearers), the speaker cannot use any referential expressions during it that depend on the accessibility of entities from the first discourse. Since the boundary between the focus space for D1 and the one for D2 is impenetrable, if D2 were to include an utterance such as "put them away," the pronoun would have to refer deictically, and not anaphorically, to the groceries.

In this sample discourse, however, D1 is resumed almost immediately. The

---

[17]This boundary is clearly atypical of stacks. It suggests that ultimately the stack model is not quite what is needed. What structure should replace the stack remains unclear to us.

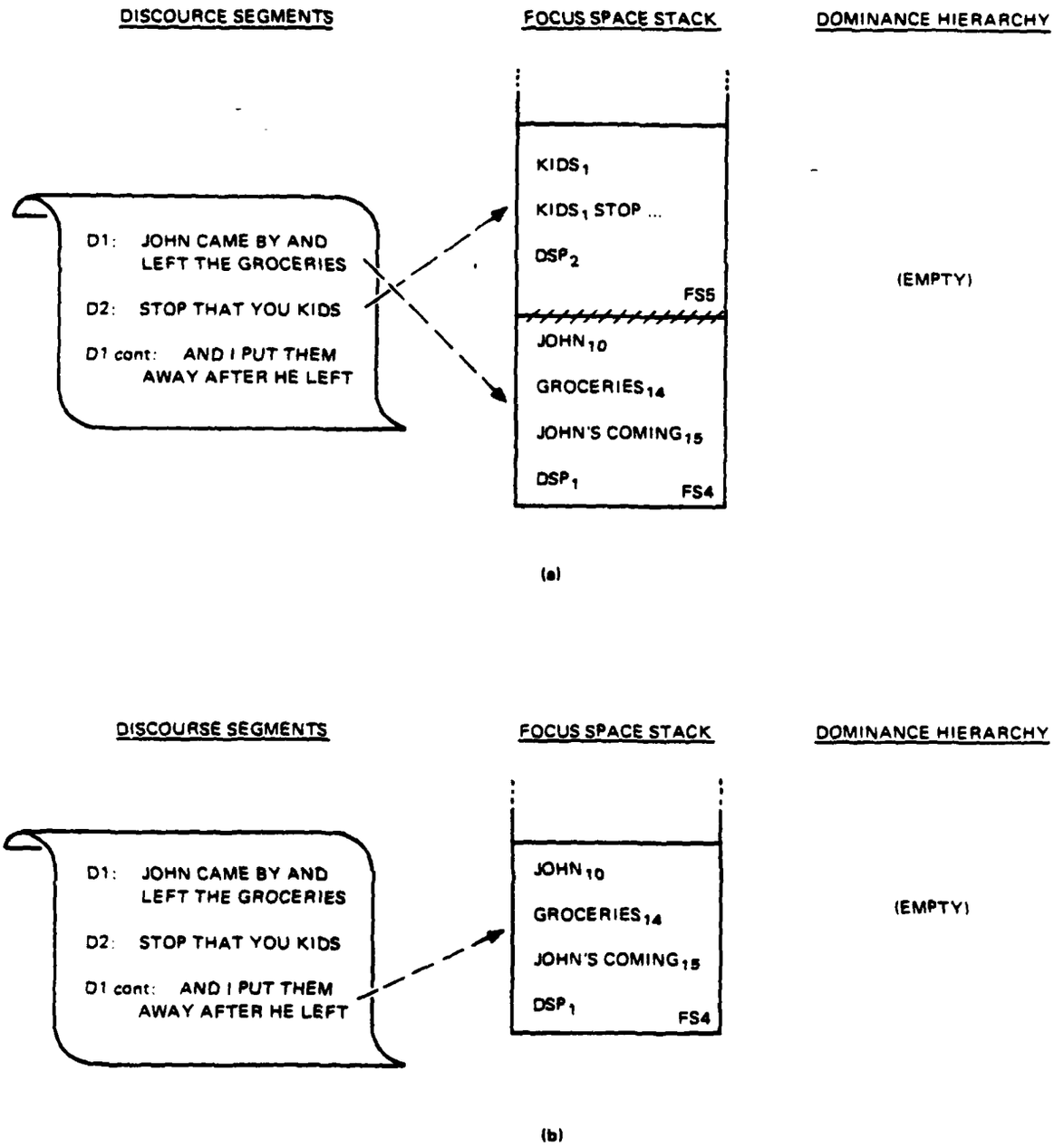


Figure 7: The Structures of a True Interruption

pronoun "them" in "and I put them away" cannot refer to the children[18](the focus space for D2 has been popped from the stack), but only to the groceries. For this to be clear to the OCP, the ICP must indicate a return to D1 explicitly. The linguistic indicators are the change of mood to imperative, and the use of the vocative [29]. Two other indicators of the "stop that" interruption are assumed to have been present at the time of the discourse--a change of intonation (imagine a slightly shrill tone of command with an undercurrent of annoyance) and a shift of gaze (toward and then away from the kids). It is also possible that the type of pause present in such cases is evidence of the interruption, but further research is needed to establish whether this is indeed the case.

In contrast to previous accounts, we are not forced to integrate these two discourses into a single grammatical structure, or to answer questions about the specific relationship between segments D2 and D1, as in Reichman's model [33]. Instead, the intuition that readers have of an embedding in the discourse structure is captured in the attentional state by the stacking of focus spaces. In addition, a reader's intuitive impression of the distinctness of the two segments is captured in their different intentional (DP/DSP) structures.

6.3 Type 2: Flashbacks and Filling in Missing Pieces

Sometimes an ICP interrupts the flow of discussion because some purposes, propositions, or entities need to be brought into the discourse but have not been: the ICP forgot to include those entities first, and so must now go back and fill in the missing information. A flashback segment occurs at that point in the discourse. The flashback is defined as a segment whose DSP satisfaction-precedes the interrupted segment and is dominated by some other segment's DSP. Hence, it is a specialization of the weak definition of interruptions. This type of interruption differs from true interruptions both intentionally and linguistically: the DSP for the flashback bears some relationship to the DP for the whole discourse. The linguistic indicator of the flashback typically includes a comment about something going wrong. In addition the

---

[18]Because this is so clearly the case on other grounds, the segment boundary is obvious even to a reader after the fact.

audience always remains the same whereas it may change for a true interruption (as in the example of the previous section).

In the example below, taken from Sidner [40], the ICP is instructing a mock-up system (mimicked by a person) about how to define and display certain information in a particular knowledge-representation language. Again the interruption is indicated by italics.

OK. Now how do I say that Bill is

Whoops I forgot about ABC.
I need an individual concept for the company ABC
...[remainder of discourse segment on ABC]...

Now back to Bill. How do I say that Bill is an employee of ABC?

The DP for the larger discourse from which this sequence was taken is to provide information about various companies (including ABC) and their employees. The outer segment in this example--D_{Bill}--has a DSP--DSP_{Bill}--to tell about Bill, while the inner segment--D_{ABC}--has a DSP--DSP_{ABC}--to convey certain information about ABC. Because of the nature of the information being told, there is order in the final structure of the DP/DSPs: information about ABC must be conveyed before all of the information about Bill can be. The ICP in this instance does not realize this constraint until after he begins. The "flashback" interruption allows him to satisfy DSP_{ABC} while suspending satisfaction of DSP_{Bill} (which he then resumes). Hence, there is an intentional structure rooted at DP and with DSP_{ABC} and DSP_{Bill} as *ordered* sister nodes. The following three relationships hold between the different DSPs:[19]

DP DOM DSP_{ABC}
DP DOM DSP_{Bill}
DSP_{ABC} SP DSP_{Bill}

This kind of interruption is distinct from a true interruption because there is a connection, although indirect, between the DSPs for the two segments. Furthermore, the linguistic markers of the start of the interruption signify that there is a

[19]From just the fragment presented, all that can be determined is that the two dominates relationships are domination but not direct domination.

precedence relation between these DSPs (and hence that the correction is necessary). Flashbacks are also distinct from normally embedded discourses by the precedence relationship between the DSPs for the two segments and the order in which the segments occur.

The available linguistic data permit three possible attentional states as appropriate models for flashback-type interruptions: one is identical to the state that would ensue if the flashback segment were a normally embedded segment, the second resembles the model of a true interruption, and the third differs from the others by requiring an auxiliary stack. An example of the stack for a normally embedded sequence is given in Section 8.

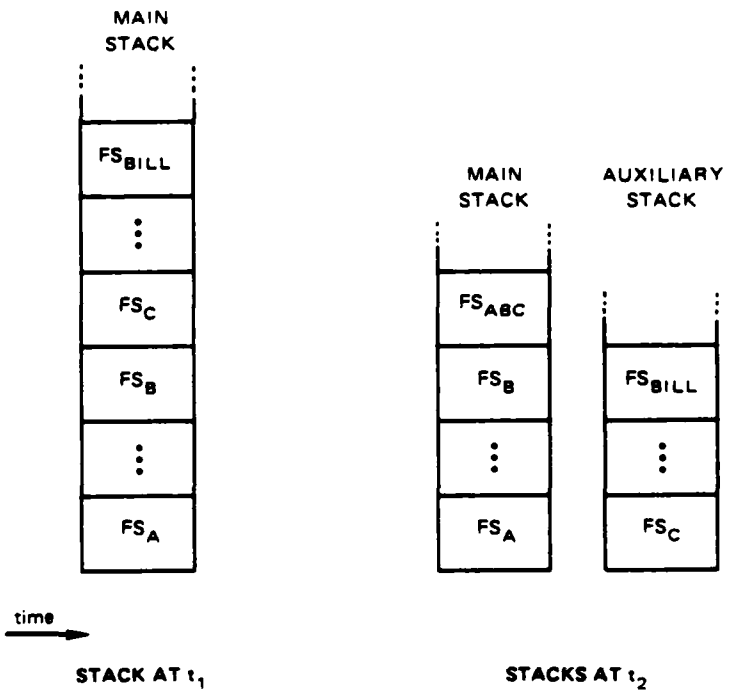


Figure 8: The Auxiliary Stack Model for Flashbacks

Figure 8 illustrates the last possibility. The focus space for the flashback--FS_{ABC}-- pushed onto the stack after an appropriate number of spaces, including the focus space for the outer segment--FS_{Bill}, have been popped from the main stack and pushed onto an auxiliary stack. All of the entities in the focus spaces remaining on the main stack are normally accessible for reference, but none of those on the auxiliary stack are. In the example in the figure, entities in the spaces from FS_A to FS_B are accessible as well (though less salient than) those in space FS_{ABC}. Evidence for this kind of stack behavior could come from discourses in which phrases in the segment about ABC could refer to entities represented in FS_B, but not to those in FS_{Bill} or FS_C. After an explicit indication that there is a return to DSP_{Bill} (e.g., the "Now back to Bill" used in this example), any focus spaces left on the stack from the flashback are popped off, and all spaces on the auxiliary stack (including FS_{Bill}) are returned to the main stack. Note, however, that this model does not preclude the possibility of a return to some space between FS_A and FS_C before popping the auxiliary stack. Whether such a return is possible remains an open question.

The auxiliary stack model differs from the other two by the references it allows and by the spaces that can be popped to. Given the initial configuration in Figure 8, if the segment with DSP DSP_{ABC} were normally embedded, FS_{ABC} would just be added to the top of the stack. If it were a true interruption, the space would also be added to the stack, but with an impenetrable boundary between it and FS_{Bill}. In the normal stack model, entities in the spaces lower in the stack would be accessible; in the true interruption they would not. In either of these two models, however, FS_{Bill} would be the space returned to first. The auxiliary stack model is obviously more complicated than the other two alternatives. Whether it (or some equivalent alternative) is necessary depends on facts of discourse behavior that have not yet been determined.

6.4 Type 3: Digressions

The third type of interruption, which we call a digression, is defined as a strong interruption that contains a reference to some entity that is salient in both the interruption and the interrupted segment. For example, if while discussing Bill's role in company ABC, one conversational participant interrupts with, "Speaking of Bill, that reminds me, he came to dinner last week," Bill remains salient, but the DP changes.

Digressions commonly begin with phrases such as "speaking of John" or "that reminds me."

In the processing of digressions, the discourse-level intention of the digression forms the base of a separate intentional structure, just as in the case of true interruptions. A new focus space is formed and pushed onto the stack, but it contains at least one--and possibly other--entities from the interrupted segment's focus space. Like the flashback-type interruption, the digression must usually be closed with an explicit utterance such as "getting back to ABC..."

6.5 Noninterruptions--"Semantic Returns"

One case of discourse behavior that we must distinguish comprises the so-called "semantic returns" observed by Reichman [32] and discussed by Polanyi and Scha [29]. In all the interruptions we have considered so far, the stack must be popped when the interruption is over and the interrupted discourse is resumed. The focus space for the interrupted segment is "returned to." In the case of semantic returns, entities and DSPs that were salient during a discourse in the past are taken up once again, but are explicitly reintroduced. For example, suppose that yesterday two people discussed how badly Jack was behaving at the party; then today one of them says "Remember our discussion about Jack at the party? Well, a lot of other people thought he acted just as badly as we thought he did." The utterances today recall, or return to, yesterday's conversation to help satisfy the intention that more be said about Jack's poor behavior.

Anything that can be talked about once can be talked about again. However, if there is no focus space on the stack corresponding to the segment and DSP being discussed further, then, as Polanyi and Scha [29] point out, there is no popping of the stack. There need not be any discourse underway when a semantic return occurs; in such cases, the focus stack will be empty. Thus, unlike the returns that follow normal interruptions, semantic returns involve a push onto the stack of a new space containing, among other things, representations of the reintroduced entities.

The separation of attentional state from intentional structure makes clear not only what is occurring in such cases, but also the intuitions underlying the term

"semantic return." In reintroducing some entities from a previous discourse, conversational participants are establishing some connection between the DSP of the new segment and the intentional structure of the original discourse. It is not a return to a previous focus space because the focus space for the original discourse is gone from the stack, and the items to be referred to must be re-established explicitly. For example, the initial reference to Jack in the preceding example cannot be accomplished with a pronoun; with no prior mention of Jack in the current discussion, one cannot say, "Remember our discussion about him at the party." The intuitive impression of a return in the strict sense is only a return to a previous intentional structure.

7 Application of the Theory: Clue Words

Both attentional state and intentional structure change during a discourse. ICPs rarely change attention by directly and explicitly referring to attentional state (e.g., using the phrase "Now let's turn our attention to..."). Likewise, discourses only occasionally include an explicit reference to a change in purpose (e.g., with an utterance such as "Now I want to explain the theory of dynamic programming"). More typically, ICPs employ indirect means of indicating that a change is coming and what kind of change it is. Clue words and phrases provide abbreviated, indirect means of indicating these changes.

In all discourse changes, the ICP must provide information that allows the OCP to determine all of the following: (1) that a change of attention is imminent; (2) whether the change returns to a previous focus space or creates a new one; (3) how the intention is related to other intentions; (4) what precedence relationships, if any, are relevant; (5) what intention is entering into focus. Clue phrases can pack in all of this information, except for (5). In this section, we will explore the predictions of our discourse structure theory about different uses of these phrases and the explanations the theory offers for their various roles.

We will use the configuration of attentional state and intentional structure illustrated in Figure 9 as the starting point of our analysis. In the initial configuration, the focus space stack has a space with DSP X at the bottom and

another space with DSP A at the top. The intentional structure includes the information that X dominates A. From this initial configuration, a wide variety of moves may be made. We will examine several changes and the clue words and phrases that can indicate each of them. Because these phrases and words in isolation may ambiguously play either discourse or other functional roles, we will also discuss the other uses whenever appropriate.

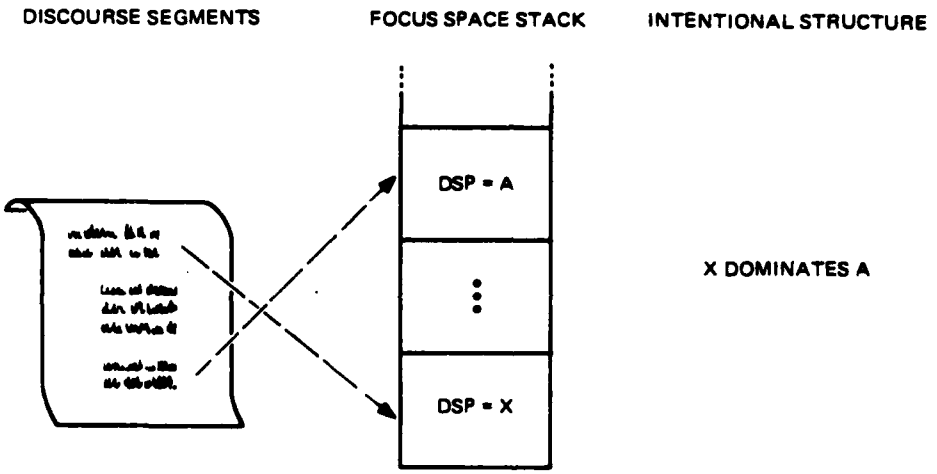


Figure 9: An Initial Discourse Structure Configuration

First, consider what happens when the ICP shifts to a new DSP, B, that is dominated by A (and correspondingly by X). The dominance relationship between A and B becomes part of the intentional structure. In addition, the change in DSP results in a change to the focus stack. The focus stack models this change, which we will call *new dominance*, by a having new space pushed onto the stack with B as the DSP of that space (as illustrated in figure 10). The space containing A is salient, but less so than the space with B. Clue phrase(s) to signal this case, and only this one, must

communicate two pieces of information: that there is a change to some new purpose (resulting in a new focus space being created in the attentional state model rather than a return to one on the stack) and that the new purpose (DSP B) is dominated by DSP A. Typical clue words for this kind of change are "for example" and "to wit."

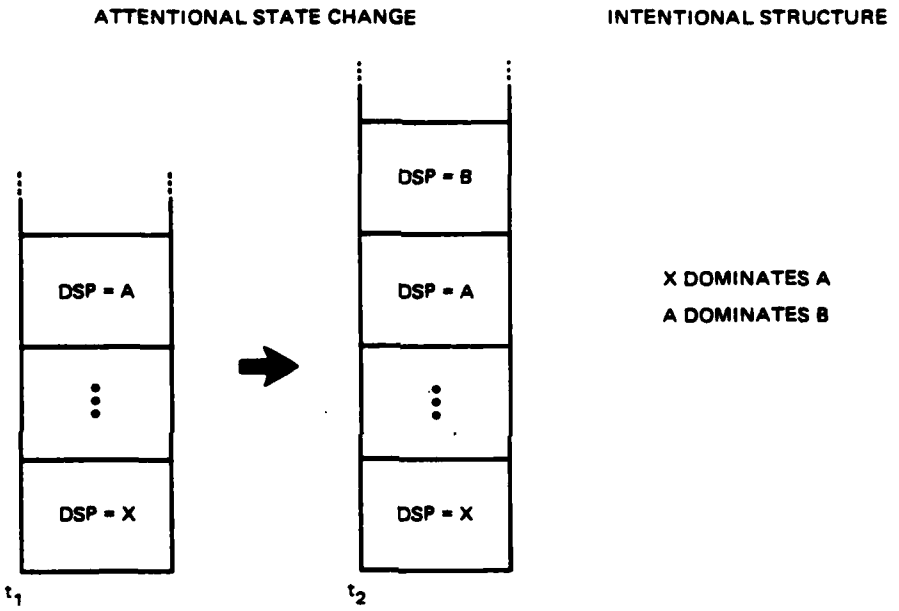


Figure 10: Attentional and Intentional Structures for a New Subsegment

Clue words can also exhibit the existence of a satisfaction-precedence relationship. If B is to be the first in a list of DSPs dominated by A, then words such as "first" and "in the first place" can be used to communicate this fact. Later in the discourse, clue words like "second," "third," and "finally" can be used to indicate DSPs that are dominated by A and satisfaction-preceded by B. In these cases, the focus space containing B would be popped from the stack and the new focus space inserted above the one containing A.

There are three other kinds of discourse segments that change the intentional

structure with a resulting push of new focus spaces onto the stack: the true-interruption, where B is not dominated by A; the flashback, where B satisfaction-precedes A; and the digression, where B is not dominated by A, but some entity is carried over to the new focus space.

One would expect that there might be clue words that would distinguish among all four of these kinds of changes. Just that is so. There are clue words or phrases that announce one and only one kind of change. The clue words mentioned above for new dominance are never used for the other three kinds of discourse pushes. The clue phrases for true-interruptions express the intention to interrupt (e.g. "Excuse me a minute, " or "I must interrupt") while the typical clue phrase for flashbacks (e.g. "Oops, I forgot about ...") indicates that something is out of order. The typical opening clue phrases of the digression mention the entity that is being carried forward (e.g. "Speaking of John ..." or "Did you hear about John?").

Clue phrases can also exhibit the satisfaction of a DSP, and hence the completion of a discourse segment. The completion of a segment causes expectations of a new piece of intentional structure and of pops of the stack. There are many means of linguistically marking completions. In texts, paragraph and chapter boundaries and explicit comments (e.g. "The End") are common. In conversations, completion can be indicated either with clue words such as "fine" or "OK"[20] or with more explicit references to the satisfaction of the intention (e.g., "That's all for point 2," or "The ayes have it.").

Most clue phrases that communicate changes to attentional state announce pops of the focus stack, but at least one clue phrase can be construed to indicate a push, namely, "That reminds me." By itself, this phrase does not specify any particular change in intentional structure, but merely shows that there will be a new DSP. Since this is equivalent to indicating that a new focus space is to be pushed onto the stack, this clue phrase is best seen as conveying attentional information.

---

[20] "OK" is many ways ambiguous. It may also mean (at least) "I heard what you said," "I heard and intend to do what you intend me to intend," "I am done what I undertook to do," or "I approve what you are about to do."

Clue phrases that indicate pops of the stack include "but anyway," "anyway," "in any case," and "now back to..." When the current focus space is popped from the stack, a space already on the stack becomes most salient. From the configuration in Figure 9, the space with A is popped from the stack, perhaps with others, and another space on the stack becomes the top of the stack. Popping back changes the stack without creating a new DSP, or a dominance or satisfaction-precedence relationship. The pop entails a return to an old DSP; no change is effected in the intentional structure.

There are clue phrases, such as "now" and "next," which signal a change of attentional state, but do not distinguish between the creation of a new focus space and the return to an old one. These words can be used for either move. For example, in a task-oriented discourse during which some task has been mentioned but put aside to ask a question, the use of "now" indicates a change of focus. The utterance following "now," however, will either return the discussion to the deferred task or will introduce some new task for consideration.

Note, finally, that a pop of the focus stack may be achieved without the use of clue phrases as in the following fragment of a task-oriented dialogue [15]:

A: One bolt is stuck. I'm trying to use both the pliers and the wrench to get it unstuck, but I haven't had much luck.
E: Don't use pliers. Show me what you are doing.
A: I'm pointing at the bolts.
E: Show me the 1/2" combination wrench, please.
A: OK.
E: Good, now show me the 1/2" box wrench.
A: I already got it loosened.

The last utterance in this fragment returns the discourse to the discussion of the unstuck bolt. The pop can be inferred only from the content of the main portion of the utterance. The pronoun (or, more accurately, the fact that it cannot be referring to the wrench) is a clue that a pop is needed, but only the reference to the loosening action allows the OCP to recognize to which discourse segment this utterance belongs, as discussed by Sidner [38] and Robinson [35].

The cases listed here do not exhaust the changes in focus spaces and in the dominance hierarchy that can be represented--nor have we furnished a set of rules

that specify when clue phrases are necessary. Additional cases, especially special subcases of these, may be possible. When discourse is viewed in terms of intentional structure and attentional state, it is clearer just what kinds of information linguistic expressions and intonation convey to the hearer about the discourse structure. Furthermore, it is clear that linguistic expressions can function as clue words, as well as sentential connections; they can tell the hearer about changes in the discourse structure and be carriers of discourse, rather than sentence-level semantic, meaning.

8 Application of the Theory: Referring Expressions

Let us now return briefly to the task-oriented dialogue in Section 3.2, to illustrate the effect of discourse segmentation and the attentional state on the use of referring expressions. The phrase "the screw" in (25) of that fragment is of particular interest. As Grosz [17] notes the two setscrews discussed in (3) through (18) are not focused on by either participant at the utterance of (25). Hence, those objects were not considered as possible referents for the two uses of "the screw" in (25), both of which refer to the screw in the center of the wheelpuller. (The wheelpuller has three screws, two on the arms and one in the center; hence the modifier "in the center" is essential in the initial description.)

In the current framework, these facts can be explained by the focus stack configuration when (25) is spoken. The stack will contain (in bottom-to-top order) focus spaces FS1, FS4, and FS5 for segments DS1, DS4, and DS5, respectively. In DS5 the wheelpuller is a focused entity, while in DS4 the two setscrews are in focus (because they are important parts of the flywheel). Since FS5 is used before FS4 to provide referents for reduced noun phrases, such as "the screw in the center" and "the screw," the wheelpuller's center screw will be identified as the referent.

To explain the use of pronouns in discourse, a second level of focusing plays a

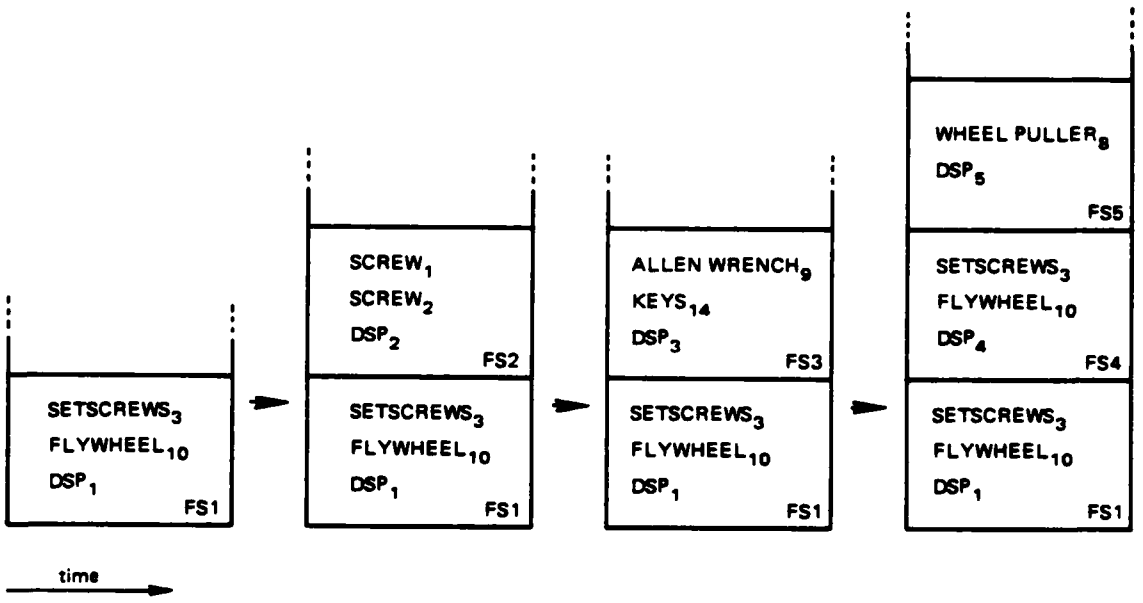


Figure 11: Focus Stack Transitions Leading up to Utterance (25)

central role in the processing of anaphoric expressions in discourse,[21] and is part of the attentional state. As described by Grosz and her collegues [19], a *backward-looking center* is associated with each utterance in a discourse segment for use in interpreting anaphoric expressions. In addition to the information described

[21] Linguistic evidence supporting the existence of these two levels includes the differential use of pronominal and nonpronominal referring expressions [16, 33]. Although Reichman [33] observes this differential use across a wide range of discourse types, she actually distinguishes four levels of focusing, corresponding to four modes of reference: pronominal, proper name, definite description, and elliptical. However, the middle two levels collapse to one as soon as the differences between the shared knowledge underlying definite descriptions and that underlying proper names is accounted for. Her fourth level arises from the mistaken classification of some clauses as elliptical.

earlier, each focus space tracks the center of the current utterance. The center distinguishes among all the focused elements the one that is central at that utterance. Centering, like focusing, is a dynamic behavior, but is a more local phenomenon; centering rules constrain the use of anaphors across the utterance boundaries within a segment.

At any moment, each focus space thus includes two distinguished items, the DSP and the current backward-looking center. The center differs crucially from the DSP because it may shift within a discourse segment (it almost always shifts across segment boundaries) whereas the DSP does not: a change in DSP underlies a segment boundary. In addition, the center is an element of the attentional state only, whereas the DSP plays a role in both the attentional and intentional components.

The existence of a center in the focus space leads to certain unanswered questions about their use at discourse segment boundaries. How is centering influenced by a discourse segment boundary? How is it affected by the DSPs and their relationships (DOM and SP)? If the DSP of a new discourse segment is a sister of the DSP that has just been satisfied (i.e., its segment has just finished), can the backward-looking center from the last utterance of the just finished segment be continued by means of a pronoun, even though the focus space is about to be popped? When a new segment is embedded relative to the current segment, must the center be expressed with a fuller noun phrase than a personal pronoun or bare demonstrative (i.e., "this" or "that")? A previous example in section 7 illustrated the use of a pronoun in an utterance that returned the discourse to an earlier segment causing the focus stack to be popped. When focus spaces are popped to return to an earlier space, what are the constraints on pronominalization?

The presence of both centers and DSPs in this theory leads us to an intriguing conjecture: that "topic" is a concept that is used ambiguously for both the DSP of a segment and the center. In the literature the concept of "topic" has appeared in many guises. In syntactic form it is used to describe the preposing of syntactic constituents in English and the "wa" marking in Japanese. Researchers have used it to describe the sentence topic (i.e., what the sentence is about [11, 37, 20]), and as a pragmatic notion [34]; others want to use the term for "discourse topic" either to mean what the discourse is about, or to be defined as those proposition(s) the ICP

provides or requests new information about (see Reinhart [34] for a review of many of the notions of aboutness and topic). It appears that many of the descriptions of sentence topic correspond (though not always) to centers, while discourse topic corresponds to the DSP of a segment or of the discourse.

9 Conclusions and Future Research

The theory of discourse structure presented in this paper is a generalization of theories of task-oriented dialogues. It differs from previous generalizations in that it carefully distinguishes three components of discourse structure: one linguistic, one intentional, and one attentional. This distinction provides an essential basis for explaining interruptions, clue words, and referring expressions.

The particular intentional structure used also differs from the analogous aspect of previous generalizations. Although, like them, it supplies the principal framework for discourse segmentation and determines structural relationships for the focusing structure (part of the attentional state), unlike its predecessors it does not depend on the special details of any single domain or type of discourse.

Although admittedly still incomplete, the theory does provide a solid basis for investigating both the structure and meaning of discourse, as well as for constructing discourse-processing systems. Several difficult research problems remain to be explored. Of these, we take the following to be of primary importance:

1. Specification of the relationship between discourse-level (DP/DSP) and utterance-level intentions
2. Identification of the information discourse participants use to recognize these intentions, and the ways in which they utilize it
3. Providing an adequate treatment of the interaction among intentions of multiple participants
4. Investigation of the effect of multiple DSPs on the theory
5. Investigation of alternative models of attentional state

Finally, the theory suggests several important conjectures. First, that a

discourse is coherent only when its discourse purpose is shared by all the participants and when each utterance of the discourse contributes to achieving this purpose, either directly or indirectly, by contributing to the satisfaction of a discourse segment purpose. Second, that the notion of "topic" is primarily an intentional notion; it is best seen as referring to the DP/DSPs. Previous discussions of the "topic" of an utterance or discourse have been confused because the term "topic" has been used to refer to alternative notions some of which are essentially syntactic, others of which are attentional (the center of an utterance), and yet others are intentional (the DSP of a segment). Finally, the theory suggests that the same intentional structure can give rise to different attentional structures through different discourses. The different attentional structures will be manifest in part because different referring expressions will be valid, and in part because different clue words and other indicators will be necessary, optional, or redundant.

Acknowledgments: We have benefited greatly from discussions with Martha Pollack, Ray Perrault, and Scott Weinstein. The paper has benefited from the comments of Brad Goodman, David Israel, Amichai Kronfeld, Martha Pollack, Ray Perrault, John Perry, Jane Robinson, Stuart Shieber, Ralph Weischedel, Scott Weinstein. Whatever errors remain are, of course, all ours.

This paper was made possible by a gift from the System Development Foundation. Support was also provided for the second author by by the Advanced Research Projects Agency of the Department of Defense and was monitored by ONR under Contract No. N00014-85-C-0079. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

References

[1] Allen, J.F.
A plan-based approach to speech act recognition.
Technical Report 131, Department of Computer Science, University of Toronto, Toronto, Canada, January, 1979.

[2] Allen, J.F., and Perrault, C.R.
Analyzing intention in dialogues.
Artificial Intelligence 15(3):143-178, 1980.

[3] Allen, J.F.
Recognizing Intentions from Natural Language Utterances.
In M. Brady and R.C. Berwick (editors), *Computational Models of Discourse*, pages 107-166. Massachusetts Institute Technology Press, 1983.

[4] Appelt, D.
Planning English Referring Expressions.
Artificial Intelligence 26 :1-33, 1985.

[5] Butterworth, B.
Hesitation and semantic planning in speech.
Journal of Psycholinguistic Research (4):75-87, 1975.

[6] Chafe, Wallace L.
The Flow of Thought and the Flow of Language.
In T. Givon (editor), *Syntax and Semantics, Vol. 12, Discourse and Syntax*, pages 159-182. Academic Press, 1979.

[7] Chafe, W.L.
Ed., The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production. Vol. 3. Advances in Discourse Processes.
Norwood, NJ: Ablex Publishing Corp.
1980

[8] Cohen, P.R. and Levesque, H.L.
Speech Acts and the Recognition of Shared Plans.
In *Proc. of the Third Biennial Conference*, pages 263-271. Canadian Society for Computational Studies of Intelligence, Canadian Society for Computational Studies of Intelligence, Victoria, B. C., May, 1980.

[9] Cohen, R.
A Computational Model for the Analysis of Arguments.
Technical Report CSRG-151, Computer Systems Research Group, University of Toronto, October, 1983.

[10] Cohen, P.R. and Levesque, H.J.
Speech Acts and Rationality.
In *Proceedings of 23rd Annual Meeting of the ACL*, pages 49-60. Assoc. for Computational Linguistics, Chicago, IL, July, 1985.

[11] Firbas, J.
On the Concept of Communicative Dynamism in the Theory of Functional Sentence Perspective.
Technical Report, Brno Studies in English 7, 12–47, 1971.

[12] Goldman, A.I.
A Theory of Human Action.
Princeton University Press, Princeton, NJ, 1970.

[13] Grice, H.P.
Utterer's Meaning and Intentions.
Philosophical Review 68(2):147–177, 1969.

[14] Grimes, J.E.
The Thread of Discourse.
Mouton, The Hague, 1975.

[15] Grosz, Barbara [Deutsch].
The Structure of Task Oriented Dialogs.
In *IEEE Symposium on Speech Recognition: Contributed Papers.* IEEE, Pittsburgh: Carnegie Mellon University Computer Science Dept., 1974.
Reprinted in L. Polanyi (ed.), The Structure of Discourse, Vol. in the Advances in Discourse Processing Series, 1986.

[16] Grosz, B.J.
Discourse Analysis.
In D. Walker (editor), *Understanding Spoken Language*, chapter IX, pages 235–268. Elsevier North–Holland, New York City, 1978.

[17] Grosz, B.J.
Focusing in Dialog.
In *Theoretical Issues in Natural Language Processing–2*, pages 96–103. The Association for Computational Linguistics, University of Illinois at Urbana–Champaign, July, 1978.

[18] Grosz, B.J.
Focusing and Description in Natural Language Dialogues.
In A. Joshi, B. Webber, I. Sag (editors), *Elements of Discourse Understanding*, pages 84–105. Cambridge University Press, 1981.

[19] Grosz, B.J., Joshi, A.K., Weinstein, S.
Providing a Unified Account of Definite Noun Phrases in Discourse.
In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics, June, 1983.

[20] Hajicova, E.
Topic and Focus.
Theoretical Linguistics 10(2/3):268–276, 1983.

[21] Hendrix, G.G.
Encoding Knowledge in Partitioned Networks.
In Nicholas V. Findler (editor), *The Representation and Use of Knowledge in Computers.* Academic Press, New York, 1979.

[22] Hobbs, J.
Coherence and Co-references.
Cognitive Science 3(1):67-82, 1979.

[23] Holmes, H.W. and Gallagher, O.
Composition and Rhetoric.
D. Appleton and Co., New York, 1917.

[24] Linde, C. and Goguen, J.
Structure of Planning Discourse.
J. Social Biol. Struct. 1:219-251, 1978.

[25] Linde, C.
Focus of Attention and the Choice of Pronouns in Discourse.
In T. Givon (editor), *Syntax and Semantics, Vol. 12 of Discourse and Syntax*, pages 337-354. Academic Press, Inc., 1979.

[26] Mann, W.C., Moore, M.A., Levin, J.A., Carlisle, J.H.
Observation Methods for Human Dialogue.
Technical Report, Information Sciences Institute, RR/75/33, Marina del Rey, CA, June, 1975.

[27] W.C. Mann and S.A. Thompson.
Relational Propositions in Discourse.
Technical Report RR-83-115, Information Sciences Institute, November, 1983.

[28] Marcus, M. P., Hindel, D., and Fleck, M. M.
D-Theory: Talking about Talking about Trees.
In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pages 129-136. The Association for Computatinal Linguistics, Cambridge, MA, June, 1983.

[29] Polanyi, L. and Scha, R.
On the Recursive Structure of Discourse.
In K. Ehlich and H. van Riemsdijk (editors), *Connectedness in Sentence, Discourse and Text*, pages 141-178. Tilburg University, 1983.

[30] Polanyi, L. and Scha, R.
A Syntactic Approach to Discourse Semantics.
In *Proceedings of Int'l. Conference on Computational Linguistics.* Stanford University, Stanford, CA, 1984.

[31] Polanyi, L. and Scha, R.J.H.
Syntactic and Semantic Aspects of Discourse Structure.
In L. Polanyi (editor), *The Structure of Discourse.* Ablex Publishing Co., Norwood, NJ, 1986.
Volume in the Advances in Discourse Processing Series.

[32] R. Reichman.
Plain-speaking: A theory and grammar of spontaneous discourse.
PhD thesis, Department of Computer Science, Harvard University, 1981.
Also, BBN Report No. 4681, Bolt Beranek and Newman Inc., Cambridge, MA.

[33] Reichman-Adar, R.
Extended Person-Machine Interface.
Artificial Intelligence 22(2):157-218, March, 1984.

[34] Reinhart, T.
Conditions on Text Coherence.
Poetics Today , 1981.

[35] Robinson, A.
Interpreting verb phrase references in dialogs.
In *Proceedings of the Third Biennial Conference of the Canadian Society for Computational Studies of Intelligence*. Victoria, May, 1980.

[36] Schank, R.C., Collins, G.C., Davis, E., Johnson, P.N., Lytinen, S., Reiser, B.J.
What's the Point?
Cognitive Science 6(3):255-275, July-September, 1982.

[37] Sgall, P., Hajicova, E. and Benesova, E.
Topic, Focus and Generative Semantics.
Technical Report, Kronberg, 1973.

[38] Sidner, C.L.
Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse.
Technical Report 537, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, June, 1979.
PhD. Thesis.

[39] Sidner, C.L., and Israel, D.J.
Recognizing intended meaning and speaker's plans.
In *Proceedings of the International Joint Conference in Artificial Intelligence*, pages 203-208. IJCAI, IJCAI, Vancouver, B.C., August, 1981.

[40] Sidner, C.L.
Protocols of Users Manipulating Visually Presented Information with Natural Language.
Technical Report 5128, Bolt Beranek and Newman Inc., September, 1982.

[41] Sidner, C.L.
What the Speaker Means: The Recognition of Speakers' Plans in Discourse.
Computers and Mathematics with Applications 9(1), 1983.
Special Issue on Computational Linguistics - Nick Cercone, guest editor.

[42] Sidner, C.L.
What the Speaker Means: The Recognition of Speakers' Plans in Discourse.
International Journal of Computers and Mathematics, Vol. 9, No. 1 , 1983.

[43] Sidner, C.L.
Plan parsing for intended response recognition in discourse.
Computational Intelligence 1(1):1-10, February, 1985.

[44] Walker, D.
Understanding Spoken Language.
Elsevier North-Holland, New York City, 1978.

[45] Wittgenstein,L.
Philosophical Investigations.
Oxford Press, 1953.

Official Distribution List

Contract N00014-85-C-0079

	Copies
Scientific Officer Head, Information Sciences Division Office of Naval Research 800 North Quincy Street Arlington, VA 22217-5000 Attn: Dr. Alan L. Meyrowitz	1
Mr. Frank Skieber Defense Contract Administration Services Region - Boston 495 Summer Street Boston, MA 02210-2184	1
Director, Naval Research Laboratory Attn: Code 2627 Washington, DC 20375	1
Defense Technical Information Center Bldg. 5 Cameron Station Alexandria, VA 22314	12

END

FILMED

2-86

DTIC